EUROPEAN PATENT OFFICE

Patent Abstracts of Japan

PUBLICATION NUMBER : 05046324
PUBLICATION DATE : 26-02-93

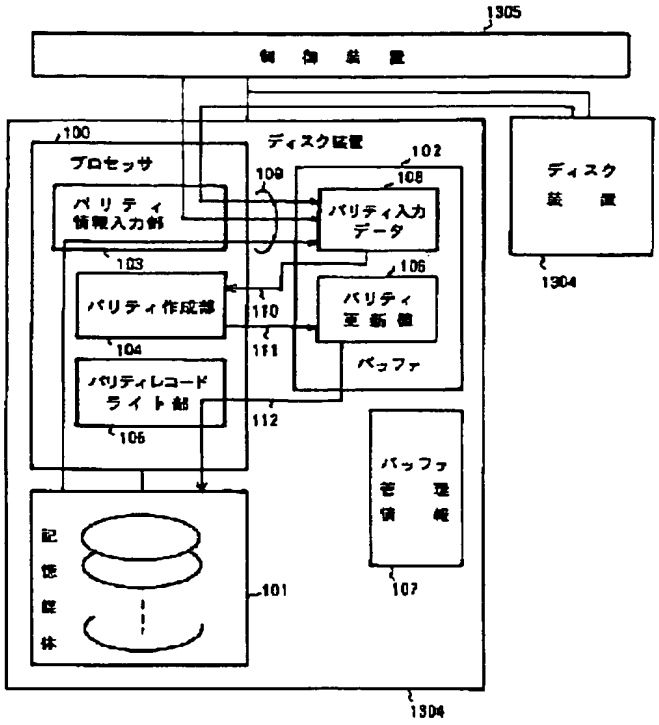
APPLICATION DATE : 20-08-91
APPLICATION NUMBER : 03207808

APPLICANT : HITACHI LTD;

INVENTOR : INOMATA HIROFUMI;

INT.CL. : G06F 3/06 G06F 11/10 G06F 12/16 G06F 12/16

TITLE : STORAGE DEVICE AND STORAGE DEVICE SYSTEM



ABSTRACT : PURPOSE: To curtail the utilization factor of a data transfer line of a controller and a disk device by dispersing a generating function of an updating value of a parity record to the disk device side, since updating of the parity record becomes overhead, comparing with a conventional processing, in the case a disk array is used.

CONSTITUTION: From a parity information input part 103 in a processor 100, a controller 1305, or a recording medium, information for generating an updating value of a parity record is inputted to a buffer. A parity generating part 104 generates the updating value of the parity record as a parity updating value 106 from the information inputted from the buffer, and stores it in the buffer. Thereafter, by a parity record write part 105, the parity updating value 106 stored in the buffer is written in a recording medium.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平5－46324

(43)公開日　平成５年(1993)２月26日

(51)Int.Cl.[5]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ	3/06	３０４　Ｅ	7165－5B		
	11/10	３２０　Ｆ	7832－5B		
	12/16	３１０　Ｊ	7629－5B		
		３２０　Ｈ	7629－5B		

審査請求　未請求　請求項の数７(全 32 頁)

(21)出願番号　特願平3－207808

(22)出願日　平成３年(1991)８月20日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　山本　彰
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　田宮　敏彦
神奈川県小田原市国府津2880番地　株式会社日立製作所小田原工場内

(72)発明者　▲高▼松　久司
神奈川県小田原市国府津2880番地　株式会社日立製作所小田原工場内

(74)代理人　弁理士　小川　勝男

最終頁に続く

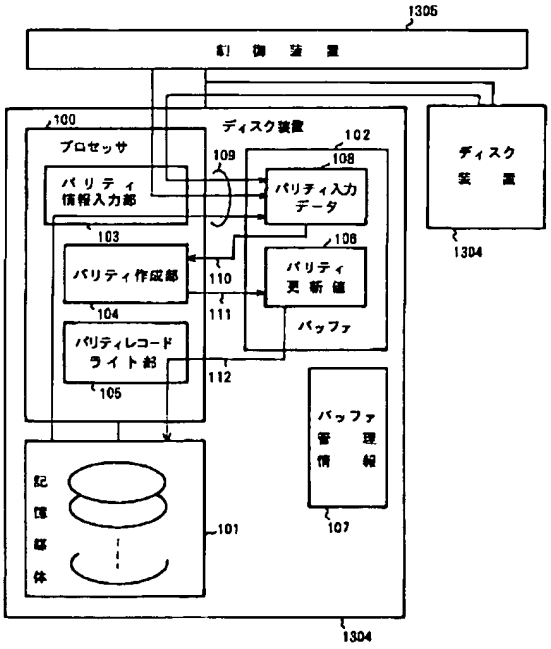
(54)【発明の名称】　記憶装置および記憶装置システム

(57)【要約】

【目的】ディスクアレイを用いた場合、パリティレコードの更新が従来処理に比較し、オーバヘッドとなる。このパリティレコードの更新値の作成機能を、ディスク装置側に分散させることにより、制御装置とディスク装置のデータ転送路の利用率を削減する。

【構成】プロセッサ内のパリティ情報入力部、制御装置、あるいは、記録媒体から、パリティレコードの更新値を作成するための情報を、バッファに入力する。パリティ作成部は、バッファに入力した情報より、パリティレコードの更新値をパリティ更新値として作成し、バッファに格納する。この後、パリティレコードライト部により、バッファに格納したパリティ更新値を記録媒体に書き込む。

図 1



【特許請求の範囲】

【請求項１】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体を含む記憶装置であって、前記冗長レコードに格納する前記冗長データを作成するための情報を、制御装置、あるいは、前記記憶媒体から受け取る手段と、前記冗長レコードに格納する前記冗長データを作成する手段を有することを特徴とする記憶装置。

【請求項２】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体を含む１つ以上の記憶装置と制御装置を含む記憶システムであって、前記制御装置が、前記データレコードの更新前の値と更新後の値から、前記冗長レコードの更新値を作成するための中間情報を作成する手段と、前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に作成した前記中間情報を送信する手段を有し、前記記憶装置が、前記制御装置から前記中間情報を受け取る手段と、前記記憶媒体から前記冗長レコードの更新前の値を読みだす手段と、前記中間情報と前記冗長レコードの更新前の値から、前記冗長レコードの更新値を作成する手段を有することを特徴とする記憶装置システム。

【請求項３】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体を含む１つ以上の記憶装置と制御装置を含む記憶システムであって、前記制御装置が、前記パリティグループ内の前記データレコードの更新値の集合を、前記パリティグループ内の前記データレコード、あるいは、前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置にブロードキャストする手段を有し、前記データレコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記パリティグループ内の前記データレコードの更新値の集合を受け取る手段と、前記パリティグループ内の前記データレコードの更新値の集合から、前記記憶媒体に書き込むべき前記データレコードの更新値を選択し、前記記憶媒体に書き込む手段と、前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記パリティグループ内の前記データレコードの更新値の集合を受け取る手段と、前記パリティグループ内の前記データレコードの更新値の集合から、前記冗長レコードの更新値を作成する手段を有することを特徴とする記憶装置システム。

【請求項４】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体を含む１つ以上の記憶装置を、接続した制御装置であって、前記制御装置が、前記データレコードの更新値を、前記データレコードを格納した前記記憶媒体を含む前記記憶装置と、前記データレコードが属する前記パリティグループ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に、ブロードキャストする手段を有し、前記データレコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記データレコードの更新値を受け取る手段と、前記データレコードの更新前の値を、前記データレコードが属する前記パリティグループ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に、送信する手段と、前記データレコードの更新値を前記記憶媒体に書き込む手段とを有し、前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記データレコードの更新値を受け取る手段と、前記データレコードを格納した前記記憶媒体を含む前記記憶装置から、前記データレコードの更新前の値を、受け取る手段と、前記記憶媒体から前記冗長レコードの更新前の値を読みだす手段と、前記データレコードの更新値と前記データレコードの更新前の値と、前記冗長レコードの更新前の値から、前記冗長レコードの更新値を作成する手段を有することを特徴とする記憶装置システム。

【請求項５】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体とバッファを含む１つ以上の記憶装置を、接続した制御装置であって、前記制御装置が、前記データレコードの更新値を、前記データレコードを格納した前記記憶媒体を含む前記記憶装置と、前記データレコードが属する前記パリティグループ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に、ブロードキャストする手段を有し、前記データレコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記データレコードの更新値を受け取る手段と、前記データレコードが属する前記パリティグル

ープ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置の要求にしたがって、前記データレコードの更新前の値を、前記データレコードが属する前記パリティグループ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に、送信する手段とを有し、前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置が、前記制御装置から前記データレコードの更新値を受け取り、バッファに格納する手段と、前記バッファに存在しなかった前記データレコードを格納した前記記憶媒体を含む前記記憶装置から、前記データレコードの更新前の値を受け取り、前記バッファに格納する手段と、前記記憶媒体から前記冗長レコードの更新前の値を読みだし、前記バッファに格納する手段と、前記バッファに格納した、前記データレコードの更新値と前記データレコードの更新前の値と、前記冗長レコードの更新前の値から、前記冗長レコードの更新値を作成する手段を有することを特徴とする記憶装置システム。

【請求項６】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体を含む記憶装置であって、制御装置からのリード要求にしたがって、前記データレコードの値を、少なくとも、前記制御装置と、前記データレコードが属する前記パリティグループ内の前記冗長レコードを格納した前記記憶媒体を含む前記記憶装置に、ブロードキャストする手段を有することを特徴とした記憶装置。

【請求項７】それぞれのレコードグループが、通常のデータを格納したｍ個（ｍ≧１）のデータレコードと前記通常データを回復するための冗長データを格納したｎ個（ｎ≧１）の冗長レコードから構成され、少なくとも１つ以上の前記レコードグループに属する１つ以上の前記データレコード、あるいは、１つ以上の前記冗長レコードを格納した記憶媒体とバッファを含む記憶装置であって、前記記憶媒体に格納した前記冗長レコードが属する前記パリティグループ内のデータレコードの値を、前記データレコードを格納した前記記憶媒体を含む記憶装置から受け取り、前記バッファに格納する手段を有することを特徴とする記憶装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、ディスクアレイ向きの高機能ディスク装置、および、ディスクアレイ向きの高機能ディスク装置と制御装置により構成される記憶装置サブシステムに関する。

【０００２】

【従来の技術】発明に最も近い公知例として、以下に示すＰａｔｔｅｒｓｏｎの論文が知られている。

【０００３】エー．シー．エム．シグモッド コンファレンス プロシーディング，1988年，6月，ページ１０９－１１６（D.Patterson,et al:A Case for Redundant Arraysof Inexpensive Disks(RAID),ACM SIGMOD conference proceeding,Chicago,IL,June1-3,1988,pp.109-116）Ｐａｔｔｅｒｓｏｎの論文は、ディスクアレイ上のデータ配置に関する技術を開示したものである。

【０００４】ディスクアレイは、ディスクシステムの高性能化、高信頼化を実現するための機構である。ディスクアレイでは、高性能化のために、物理的には複数のディスク装置を、処理装置に対しては１台のディスク装置に見せかける。一方、高信頼化のためには、データを格納したディスク装置に障害が発生した場合、データの回復を行うための冗長データを別のディスク装置に格納しておく。

【０００５】通常、ディスク装置のリード／ライト単位となるデータをレコードと呼ぶが、Ｐａｔｔｅｒｓｏｎの論文では、いくつかのレコード配置方法が提案されている。ただし、ディスクアレイを用いた場合、処理装置から見たリードライト単位であるレコードと、ディスク装置に実際に記録されるレコードとではデータ長が異なる場合がある。以下、前者を論理レコード、後者を物理レコードと呼ぶ。以下、Ｐａｔｔｅｒｓｏｎの論文で提案されているいくつかのレコード配置方法の説明を行う。

【０００６】第１の配置方法は、論理レコード、すなわち、処理装置側から見たレコードを、ディスク装置上では、ｍ個（ｍ≧１）の物理レコードに分割して格納する配置方法である。以下、この配置方法を、分割配置方法と呼ぶ。分割配置を用いた場合、１つの論理レコードをｍ台のディスク装置との間で転送できることから、見かけ上データ転送速度をｍ倍に向上させたのと同様の効果を得ることができる。次に、分割配置における冗長データの作成方法を説明する。分割配置では、論理レコードを分割したｍ個の物理レコードに対し、ｎ個（ｎ≧１）の冗長データが作成され、それぞれを、１つの物理レコード（全体でｎ個ある）としてディスク装置に格納する。以下、処理装置が直接リード／ライトするデータを格納した物理レコードをデータレコード、冗長データを格納した物理レコードをパリティレコードと呼ぶ。また、ｍ個のデータレコードとｎ個のパリティレコードから構成されるグループを、パリティグループと呼ぶ。通常、パリティグループ内のパリティレコードの数がｎ個であれば、ｎ台までのディスク装置に障害が発生してもそのパリティグループのデータは回復可能である。第２の配置方法は、処理装置から見たリード／ライト単位である論理レコードを、１つの物理レコード、すなわち、１つのデータレコードとして、ディスク装置上に格納する配置方法である。以下、これを非分割配置と呼ぶ。し

たがって、論理レコードは、データレコードと等価なる。（それぞれの物理レコードには、データレコードあるいはパリティレコードが割り当てられるため、物理レコードと論理レコードは必ずしも等価にならない。すなわち、１つの論理レコードは、１つの物理レコードではあるが、１つの物理レコードは、１つの論理レコードであるというわけではないし、パリティレコードである場合もある。）非分割配置の特長は、ディスクアレイを構成するそれぞれのディスク装置ごとにリード／ライト処理が実行可能な点である。（分割配置方法をとると、リード／ライトのために複数のディスク装置を専有する必要がある。）したがって、非分割配置をとると、ディスクアレイ内で実行できるリード／ライト処理の多重度を向上させることが可能となり、性能向上を実現できる。非分割配置でも、m個のデータレコードから、n個のパリティレコードを作成し、ディスク装置に格納される。ただし、分割配置の場合、パリティグループ内のデータレコードの集合が、処理装置から見た１つの論理レコードを形成するのに対し、非分割配置の場合、データレコードのそれぞれが、処理装置から見るとまったく独立した論理レコードとなる。

【０００７】一方、ディスクアレイではなく、一般のディスク装置のライト処理の高速化に関しては、ディスクキャッシュを利用した以下のような技術が開示されている。特開昭５５－１５７０５３では、ディスクキャッシュを有する制御装置において、ライトアフタ処理を利用してライト要求を高速化に関する技術が開示されている。具体的には、制御装置は、処理装置から受け付けたライトデータをキャッシュ内に書き込んだ段階で、ライト処理を完了させる。処理装置から受け付け、キャッシュ内に格納したデータのディスク装置への書き込みは、後から、制御装置のライトアフタ処理によって実行される。

【０００８】特開昭５９－１３５５６３では、高信頼性を保証しながらライト処理を高速化する制御装置に関する技術が開示されている。

【０００９】特開昭５９－１３５５６３では、制御装置内にキャッシュメモリ以外に不揮発性メモリを有し、処理装置から受け取ったライトデータをキャッシュメモリと不揮発性メモリに格納する。ディスク装置へのライトデータの書き込みは、制御装置が、ライトアフタ処理によって実行する。これにより、ライトアフタ処理の高信頼化を図る。一方、特開昭６０－１１４９４７では、２重書きディスク装置を制御するディスクキャッシュを有する制御装置に関する技術が開示されている。特開昭６０－１１４９４７では、制御装置は、処理装置から受け取ったライト要求に対し、一方のディスク装置とキャッシュメモリに、処理装置から受け取ったライトデータを書き込む。もう一方のディスク装置には、制御装置が、処理装置からのリード／ライト要求とは非同期に、キャッシュメモリに格納したライトデータを後から書き込む。制御装置が、処理装置からのリード／ライト要求とは非同期に、キャッシュメモリに格納したライトデータをディスク装置に後から書き込む動作をライトアフタ処理と呼ぶ。

【００１０】特開平２－３７４１８では、２重書きディスク装置をディスクキャッシュを利用した高性能化に関する技術が開示されている。

【００１１】特開平２－３７４１８でも、制御装置内にキャッシュメモリ以外に不揮発性メモリを有し、処理装置から受け取ったライトデータをキャッシュメモリと不揮発性メモリに格納する。２つのディスク装置へのライトデータの書き込みは、制御装置が、ライトアフタ処理によって実行する。

【００１２】特開平３－３７７４６では、ディスクキャッシュを有し、ライトアフタ処理を実行する制御装置において、ライトアフタ処理を効率よく実行することを目的としたディスクキャッシュ内のライトアフタデータの管理データ構造についての技術が開示されている。

【００１３】

【発明が解決しようとする課題】ディスクアレイを用いた場合、処理装置からライト処理を受け付けた時、論理レコードの内容の変化に伴い、パリティレコードの内容も変更する必要が生ずる。したがって、（ａ）ライト対象となった論理レコードの更新値の転送処理以外に、（ｂ）パリティレコードの書き込みを行う転送処理と、（ｃ）パリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理が、制御装置とディスク装置の間のデータ転送路に発生することになる。これらの転送処理は、ディスクアレイの適用により初めて生ずる処理であり、ディスクアレイ化により生ずる転送オーバヘッドとなる。ライト処理に対する制御装置とディスク装置の間のデータ転送路のデータ転送増大量は、分割配置、非分割配置それぞれの場合で異なる。以下、それぞれの場合について、具体的に説明する。

【００１４】分割配置の場合、処理装置から受け取ったライト対象とする論理レコードが、パリティグループ内のすべてのデータレコードの内容に相当することから、受け取った論理レコードの更新値から、パリティレコードが作成できる。このため、（ｃ）のパリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理は必要としない。以上より、制御装置とディスク装置の間のデータ転送路のデータ転送オーバヘッドは、（ｂ）のパリティレコードの書き込みを行う転送処理の書き込み転送量だけである。

【００１５】一方、非分散配置の場合、パリティレコードの更新値を作成するためには、（ｃ）パリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理として、以下の値の集合のうちのいずれかを得る処理が必要となる。

【００１６】（１）ライト処理が発生した論理レコード（＝データレコード）の更新前の値、および、パリティレコードの更新前の値。

【００１７】（２）ライト処理が発生した論理レコード（＝データレコード）が属するパリティグループ内の他のすべてのデータレコードの値。

【００１８】通常、（１）に示した値を得る処理の方が、オーバヘッドが小さいため、ライト処理が発生した場合、（１）に示した値を得る場合を前提にして、以下の説明を行う。（ｃ）のパリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理として、（１）に示した値を得る処理を実行すると、パリティレコードを１つしか設けない場合（ｎ＝１）でも、ライト処理が発生した論理レコード（＝データレコード）の更新前の値、および、パリティレコードの更新前の値と２回の転送が発生する。これ以外に、（ａ）ライト対象となった論理レコードの更新値の転送処理と、（ｂ）パリティレコードの書き込みを行う転送処理が１回ずつ発生するため、制御装置とディスク装置のデータ転送回数は、計４回発生することになる。ディスクアレイを用いない場合、ライト処理では、発生するデータ転送処理は、（ａ）ライト対象となった論理レコードの更新値の転送処理の１回であるため、非分散配置のディスクアレイでは、制御装置とディスク装置の間のデータ転送量は、従来の４倍に増大する。

【００１９】以上より、ディスクアレイの適用に伴い、制御装置が、処理装置との間で直接実行する転送処理に割り当てられる転送スループットは、上記の転送オーバヘッドの分だけ低下することになる。

【００２０】本発明の目的は、パリティレコードの操作に起因する、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドの増大を極力抑えることである。

【００２１】

【課題を解決するための手段】以上述べた課題に対する本発明の目的を、いかに達成するかについて以下に述べる。

【００２２】本発明では、パリティレコードの操作に起因する、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを抑えることを目的とした基本機能として、ディスク装置にパリティレコード作成機能を設ける。ただし、単純に、ディスク装置にパリティレコード作成機能を設けた場合、非分散配置では、転送オーバヘッドを削減できるが、分散配置では、転送オーバヘッドを削減できない。以下、転送オーバヘッドの最も少ないパリティレコードの数が１つの場合を例にして、その理由を説明する。

【００２３】すでに述べたように、従来の動作では、非分散配置では、処理装置からのライト要求に伴い、データレコードの書き込み以外に、データレコードとパリティレコードの更新前の値の読みだし、パリティレコードの書き込みという、従来の４倍の転送が発生する。一方、ディスク装置側にパリティ作成機能を設けた場合、制御装置側では、データレコードの更新前後の値からパリティレコードを作成するための中間値を作成し、これをディスク装置に転送する。中間値は、データレコードの更新前後の値の排他論理輪から作成される。この中間値と更新前のパリティレコードの値から更新後のパリティレコードが作成される。ディスク装置側では、制御装置から受け取ったパリティレコードを作成するための中間値と、記録媒体から読みだしたパリティレコードの更新前の値から、パリティレコードの更新値を作成し、記録媒体に書き込む。以上のような動作では、制御装置とディスク装置の間の転送処理は、データレコードの読みだし、書き込み、および、パリティレコードを作成するための中間値の転送となり、ディスクアレイ適用以前の３倍までの転送量に抑えることができる。

【００２４】しかし、分割配置の場合には、処理装置から受け取ったライト対象とする論理レコードの内容から、パリティレコードが作成できるため、制御装置が、パリティレコードを作成し、送信するのが最も効率的となる。したがって、ディスク装置側のパリティ作成機能を有効に活かすことができない。

【００２５】さらに、本発明では、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを抑えるために、基本機能であるディスク装置に設けたパリティレコード作成機能に、制御装置とディスク装置の間のブロードキャスト転送機能組み合わせて利用する。この場合には、分散配置でも、非分散配置でも、転送オーバヘッドを削減できる。以下、具体的な説明を行う。

【００２６】まず、分散配置の場合について説明する。この場合、制御装置は、パリティグループに属する全ディスク装置に、論理レコードを分割せず、そのままの形で、ブロードキャストする。この時、論理レコードを受け取るパリティグループ内のディスク装置は、論理レコードの一部をデータレコードとして記録すべきディスク装置と、論理レコードに対応するパリティレコードを記録すべきディスク装置に分類できる。そのディスク装置が、データレコードを書き込むべきディスク装置であれば、論理レコードの中から自装置で書き込むべき部分をデータレコードとして取りだし、記録媒体に記録する。一方、そのディスク装置が、パリティレコードを書き込むべきディスク装置であれば、論理レコードからパリティレコードを作成して、記録媒体に記録する。

【００２７】以上の機構では、制御装置が転送するのは論理レコードのみである。したがって、ディスクアレイの適用に伴い生ずる制御装置とディスク装置の間のデータ転送路の転送オーバヘッドをなくすことが可能となる。

【００２８】次に、非分散配置の場合について説明する。この場合も、制御装置は、パリティグループに属す

る全ディスク装置に、論理レコードを分割せず、そのままの形で、ブロードキャストする。この場合、論理レコードを受け取ったパリティグループ内のディスク装置は、論理レコード（＝データレコード）を記録媒体に記録すべきディスク装置、何も処理をしなくてよいディスク装置、および、論理レコードに対応するパリティレコードを記録すべきディスク装置に分類できる。パリティレコードを記録すべきディスク装置は、まず、論理レコードの更新前の値を、論理レコードを記録しているディスク装置から、転送してもらう。さらに、パリティレコードの更新前の値を記録媒体から読みだす。以上のようにして得た、パリティレコードの更新前の値、論理レコードの更新前の値、および、最初に受け取った論理レコードの更新値より、パリティレコードの更新値を作成し、記録媒体に書き込む。

【００２９】一方、論理レコード（＝データレコード）として書き込むべきディスク装置は、パリティレコードを記録すべきディスク装置に、まず、論理レコード（＝データレコード）の更新前の値を送信する。この後、制御装置から受け取った論理レコードの更新値を記録媒体に記録する。以上の動作では、制御装置とディスク装置の間のデータ転送路を用いて転送されるデータは、論理レコードの更新値と更新前の値である。したがって、この場合、制御装置とディスク装置の間のデータ転送路の転送量を、ディスクアレイ適用以前の２倍までに転送量を抑えることができる。

【００３０】

【作用】以下、本発明の作用について述べる。ただし、以下の説明では、転送オーバヘッドの最も少ないパリティレコードの数が１つの場合を前提とする。

【００３１】まず、非分散配置のディスクアレイに、基本機能であるディスク装置に設けたパリティレコード作成機能を適用した場合の作用について説明する。非分散配置では、本発明を適用しない場合、すでに述べたように、パリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理として、ライト処理が発生した論理レコード（＝データレコード）の更新前の値、および、パリティレコードの更新前の値を転送すると、計４回のデータ転送処理が、制御装置とディスク装置の間で発生する。

【００３２】一方、ディスク装置側にパリティ作成機能を設けた場合、制御装置側では、論理レコード（＝データレコード）の更新前と後の値からパリティレコードを作成するための中間値を作成し、これをディスク装置に転送する。ディスク装置側では、制御装置から受け取ったパリティレコードを作成するための中間値と、記録媒体から読みだしたパリティレコードの更新前の値から、パリティレコードの更新値を作成し、記録媒体に書き込む。したがって、制御装置とディスク装置の間で実行されるデータ転送処理は、データレコードの更新前の値、データレコードの更新値、および、パリティレコードの中間値の転送処理と、計３回の処理となる。本発明適用以前の転送回数は４回であるため、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを削減するという目的を達成することができる。

【００３３】次に、分散配置のディスクアレイに、基本機能であるディスク装置に設けたパリティレコード作成機能に、制御装置とディスク装置の間のブロードキャスト転送機能を組合せて適用した場合の作用について説明する。分散配置の場合、以上の機能を適用しないと、すでに述べたように、制御装置とディスク装置の間には、パリティレコードの転送がオーバヘッドとして発生することになる。以上の機能を適用した場合、制御装置は、パリティグループに属する全ディスク装置に、論理レコードを分割せず、そのままの形で、ブロードキャストする。論理レコードを受け取った各ディスク装置は、以下のような処理を行う。そのディスク装置が、データレコードを書き込むべきディスク装置であれば、論理レコードの中から自装置で書き込むべき部分をデータレコードとして取りだし、記録媒体に記録する。一方、そのディスク装置が、パリティレコードを書き込むべきディスク装置であれば、論理レコードからパリティレコードを作成して、記録媒体に記録する。したがって、制御装置が転送するのは論理レコードのみで、パリティレコードの転送は必要なくなるため、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを削減するという目的を達成することができる。

【００３４】最後に、基本機能であるディスク装置に設けたパリティレコード作成機能に、制御装置とディスク装置の間のブロードキャスト転送機能を組合せ、非分散配置を適用した場合の作用について説明する。

【００３５】この場合も、制御装置は、パリティグループに属する全ディスク装置に、論理レコードを分割せず、そのままの形で、ブロードキャストする。論理レコードの更新値を受け取ると、パリティレコードを記録すべきディスク装置は、まず、論理レコード（＝データレコード）の更新前の値を、論理レコードを記録しているディスク装置から、転送してもらう。さらに、パリティレコードの更新前の値を記録媒体から読みだす。以上のようにして得た、パリティレコードの更新前の値、論理レコードの更新前の値、最初に受け取った論理レコードの更新値より、パリティレコードの更新値を作成し、記録媒体に書き込む。

【００３６】一方、論理レコード（＝データレコード）を書き込むべきディスク装置は、パリティレコードを記録すべきディスク装置に、まず、論理レコードの更新前の値を送信する。この後、制御装置から受け取った論理レコードの更新値を記録媒体に記録する。以上の動作では、制御装置とディスク装置の間で実行されるデータ転送処理は、データレコードの更新前の値、および、デー

タレコードの更新値と、計２回の処理となる。分散配置の場合、本発明適用以前の転送回数は４回であるため、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを削減するという目的を達成することができる。

【００３７】

【実施例】以下、本発明の実施例を説明する。

【００３８】まず、各実施例に共通する内容について説明する。図１３は、本発明の対象となる計算機システムの構成である。計算機システムは、処理装置１３００、制御装置１３０５、および、１台以上のディスク装置１３０４により構成する。処理装置１３００は、ＣＰＵ１３０１、主記憶１３０２、および、チャネル１３０３により構成される場合があってもよい。制御装置１３０５は、処理装置１３００からのリード／ライト要求にしたがって、処理装置１３００と、ディスク装置１３０４の間で、転送処理を実行する。制御装置バッファ１３１０は、制御装置１３０５がリードライトするデータを、一時的に蓄えるバッファである。ただし、図２８に示したように、制御装置内に２つ以上のディレクタ１３０７を含み、それぞれのディレクタ１３０７が処理装置１３００からリード／ライト要求を受け付け、リード／ライト動作を実行するような構成でも本発明は有効である。

【００３９】図１４は、本発明の対象となる別の計算機システムの構成である。図１３に示した構成との差異は、制御装置１３０５が、キャッシュメモリ１３０８、ディレクトリ１３０９、不揮発性メモリ１４００、および、不揮発性メモリ管理情報１４０１を含む点である。キャッシュメモリ（以下、単にキャッシュと略す。）１３０８は、ディスク装置１３０４の中のアクセス頻度の高いデータをロードしておく。ディレクトリ１３０９には、キャッシュ１３０８の管理情報を格納する。不揮発性メモリ１４００は、不揮発の媒体であり、キャッシュ１３０８と同様に、ディスク装置１３０４の中のアクセス頻度の高いデータをロードしておく。不揮発性メモリ管理情報１４０１も不揮発の媒体であり、不揮発性メモリ１４００の管理情報を格納する。この場合、制御装置１３０５は、処理装置１３００からのリード／ライト要求とは、非同期に、ディスク装置２０５とキャッシュ１３０８との間で、リード／ライト動作を実行する。通常、処理装置１３００がディスク装置との間で、リード／ライトするデータの単位は、レコードと呼ばれる。ディスクアレイを適用しない場合、処理装置１３００から見たレコードと、ディスク装置１３０４上に格納されたレコードは、等しい。一方、ディスクアレイを適用した場合、処理装置１３００から見たレコードと、ディスク装置１３０４上に格納されたレコードが、ディスクアレイのレコード配置によって異なる場合がある。以下、ディスクアレイを適用した場合のデータの記録形式について説明する。

【００４０】次に、図１５、図１６を用いて、ディスクアレイの適用した場合のディスク装置１３０４上の記録形式について説明する。

【００４１】図１５に示すように、ディスク装置１３０４と制御装置１３０５の間で、リード／ライトされる単位、すなわち、ディスク装置１３０４に記録されている単位を、物理レコード１５０２と呼ぶ。本発明においては、ディスク装置１３０４上に格納されている物理レコード１５０２には、データレコード１５００とパリティレコード１５０１の２種類が存在する。データレコード１５００は、処理装置１３００が、直接リード／ライトするデータを格納した物理レコード１５０２である。一方、パリティレコード１５０１は、ディスク装置１３０４に障害が発生し、データレコード１５０１の内容が消失した時、その消失した内容を回復する処理に用いるレコードである。この場合、データレコード１５００の値が変更されると、これに対応して、パリティレコード１５０１の内容も変更する必要が生ずる。

【００４２】図１６は、ディスクアレイにおけるパリティグループ１６００の構成である。ディスク装置ａ１６０１からディスク装置ｄ１６０４までのｍ個のディスク装置１３０４上に、それぞれ対応するデータレコード１５００が格納されている。これらのｍ個のデータレコード１５００から、ｎ個のパリティレコード１５０１が作成され、それぞれ対応するディスク装置ｅ１６０５からディスク装置ｆ１６０６に格納される。したがって、図１６では、ｍ個のデータレコード１５００とｎ個のパリティレコード１５０１から、パリティグループ１６００が構成されている。一般に、ｎ個のパリティレコード１５０１を含むパリティグループ１６００においては、そのパリティグループ１６００内のレコード１５０２が格納されているｍ＋ｎ個ディスク装置のうち、ｎ台のディスク装置１３０４が故障しても、パリティグループ１６００内のすべてのレコード１５０２の内容を回復することができる。以上より、ディスクアレイを適用すると、ディスク装置１３０４の高信頼化が実現できる。

【００４３】図１６のパリティグループ１６００においては、ディスク装置ａ１６０１からディスク装置ｃ１６０３にデータレコード１５００が、ディスク装置ｄ１６０４からディスク装置ｅ１６０５に格納されている。ただし、ディスク装置ａ１６０１からディスク装置ｃ１６０３までのディスク装置１３０４上に格納するすべてのレコード１５０２を、データレコード１５００とする必要はない。同様に、ディスク装置ｄ１６０４からディスク装置ｅ１６０５までのディスク装置１３０４上のすべてのレコード１５０２が、パリティレコード１５００というわけではない。

【００４４】また、図１６においては、パリティグループ１６００が、ディスク装置ａ１６０１からディスク装置ｅ上に作成されているが、パリティグループ１６００

が作成されるディスク装置１３０４の集合は、各パリティグループ１６００が異なったディスク装置１３０４上に存在してもよい。例えば、別のパリティグループ１６００がディスク装置ｂ１６０２からディスク装置ｃ１６０３に作成されてもよい。同様に、パリティグループ１６００を構成するレコード１５０２の数もｍ＋ｎ個には限定されない。

【００４５】ただし、以下の実施例においては、簡略化のため、パリティグループ１６００の構成を、図１６に示したように、ｍ個のデータレコード１５００とｎ個のパリティレコード１５０１からなる構成とする。

【００４６】ディスクアレイを適用した場合、処理装置１３００から見たレコードと、ディスク装置１３０４上のデータレコード１５００の関係が、レコードの配置方法により、異なる場合がある。以下、図１７と図１８を用いて、２つの代表的なディスクアレイのレコード配置方法である分散配置と非分散配置の説明を行う。

【００４７】図１７は、分散配置のディスクアレイのレコード配置を表す。以下の実施例では、処理装置１３００と制御装置の間でリード／ライトされる単位、すなわち、処理装置から見たリード／ライト単位を論理レコード１７００と呼ぶ。分散配置のデイスクアレイでは、論理レコード１７００は、ｍ個に分割され、それぞれがデータレコード１５００として、ディスク装置１３０４に格納される。さらに、ｍ個の分割されたデータレコード１５００からｎ個のパリティレコード１５０１が作成され、それぞれディスク装置１３０４に格納される。したがって、分散配置の場合、１つの論理レコード１７００が、１つのパリティグループ１６００を構成する。

【００４８】分割配置を用いた場合、１つの論理レコード１７００をｍ台のディスク装置との間で転送できることから、見かけ上データ転送速度をｍ倍に向上させたのと同様の効果を得ることができる。ただし、１つのリード／ライト動作に対応して、少なくともｍ台のディスク装置１３０４を占有する必要が生ずる。

【００４９】図１８は、非分散配置のディスクアレイのレコード配置を表す。非分散配置では、１つの論理レコードがデータレコード１５００としてディスク装置１３０４に格納される。すなわち、論理レコード１７００とデータレコード１５００は１対１の対応関係をもつ，すなわち，等価である。したがって、非分散配置では、１つのパリティグループ１６００には、ｍ個の論理レコード１７００が含まれる。

【００５０】非分割配置をとった場合、論理レコード１７００のリード／ライトに１つのディスク装置１３０４だけを占有すればよいため、ディスクアレイ内で実行できるリード／ライト処理の多重度を向上させることが可能となる。

【００５１】ディスクアレイの適用により、分散配置あるいは非分割配置のいずれの場合も、論理レコード１７００のライト処理に伴い、パリティレコード１５０１の内容も書き換える必要が生ずる。したがって、ディスク装置１３０４と制御装置１３０５の間で、リード／ライトされるデータ量が、ディスクアレイの適用に伴い、従来に比較し、増大するという問題が発生する。特に、非分散配置の場合、リード／ライトされるデータ量の増加率が高い。以下、それぞれの配置をとった場合のリード／ライトデータの増加量について説明する。

【００５２】分割配置の場合、処理装置１３００から受け取ったライト対象とする論理レコード１７００を分割したｍ個のデータレコード１５００の内容から、パリティレコード１５０１が作成できる。また、論理レコード１７００、すなわち、ｍ個のデータレコード１５００に相当するライトデータ量は、従来のディスク装置でもライトされるため、ライトデータ量の増加はない。したがって、分割配置の場合、デイスクアレイにおいてリード／ライトされるデータの増分は、パリティレコードのライトデータ量だけである。

【００５３】一方、非分散配置では、処理装置１３００から受け取ったライト対象とする論理レコード１７００の内容だけからは、パリティレコード１５０１が作成できない。この場合、ライト処理が発生した論理レコード１７００の更新値以外に、例えば、以下の値の集合のうちのいずれかを得るためのリード処理を実行する必要がある。

【００５４】（１）ライト処理が発生した論理レコード１７００、すなわち、対応するデータレコード１５００の更新前の値、および、パリティレコード１５０１の更新前の値

（２）ライト処理が発生した論理レコード１７００（＝データレコード１５００）が属するパリティグループ１６００内の他のすべてのデータレコード１５００の値

通常、（１）に示した値を得る方が、オーバヘッドが小さいため、ライト処理が発生した場合、（１）に示した値を得る方法をとった場合について説明する。したがって、パリティレコード１５００を書き込むというライト処理に加えて、さらに、（１）に示した値を得るためのリード処理が、従来ディスク装置に比較し、転送オーバヘッドとなる。このため、パリティレコード１５００が１つしか設けない場合（ｎ＝１）でも、ディスクアレイ適用以前に比較し、リード／ライトされるデータ量は、４倍に増大する。

【００５５】本発明の概要は、ディスクアレイを適用した時、論理レコード１７００のライト処理に伴い発生する、ディスク装置１３０４と制御装置１３０５の間のリード／ライトデータの転送量を、ディスク装置１３０４にパリティレコード５０１の作成機能を設けることにより削減するというものである。以下、図１を用いて、その内容を説明する。

【００５６】まず、ディスク装置１３０４の構成を説明

する。ディスク装置１３０４内の記憶媒体１０１は、実際に物理レコード１５０１を記録した媒体である。プロセッサ１００は、制御装置１３０４から記憶媒体１０１上の物理レコード１５０２をリード／ライトする。本発明では、さらに、プロセッサ１００は、パリティレコード作成機能をもつ。バッファ１０２は、パリティレコード１５０１の更新値等を作成する際に利用するメモリである。バッファ管理情報１０７は、バッファ１０２上に、どのようなデータを格納したを示す管理情報である。

【００５７】ディスク装置１３０４は、制御装置１３０５の指示にしたがい、パリティレコード１５０１の更新処理を実行する。以下、その内容を説明する。プロセッサ１００内のパリティ情報入力部１０３は、制御装置１３０５、あるいは、記憶媒体１０１から、パリティレコード１５０１の更新値を作成するための情報を、パリティ入力データ１０８として、バッファ１０２に入力する（１０９）。パリティ作成部１０４は、パリティ情報入力部１０３によりバッファ１０２に格納されたパリティ入力データ１０８を入力し（１１０）、パリティレコード１５０１の更新値をパリティ更新値１０６として作成し、バッファ１０２に格納する（１１１）。この後、パリティレコードライト部１０５により、バッファ１０２に格納したパリティ更新値１０６を記憶媒体１０１に書き込む（１１２）。

【００５８】以上が、本発明の概要であるが、以下の実施例では、ディスク装置１３０４に設けるパリティレコード５０１の作成機能に関し、３種類の実施例を開示する。まず、各実施例の概要を説明する。

【００５９】第１の実施例は、単純に、ディスク装置１３０４にパリティレコード作成機能を設け、非分散配置をとったディスクアレイに関する。

【００６０】以下、第１の実施例の概要を図２を用いて説明する。すでに述べたように、非分散配置をとるディスクアレイでは、論理レコード１７００（＝データレコード１５００）の更新値、ライト処理が発生した論理レコード１７００の更新前の値、および、パリティレコード１５０１の更新前の値により、パリティレコード１５０１の更新値を作成することができる。

【００６１】ディスク装置１３０４にパリティレコード作成機能がなければ、非分散配置では、処理装置からのライト要求に伴い、制御装置１３０５とディスク装置１３０４の間で、論理レコード１７００（＝データレコード１５００）とパリティレコード１５０１の読みだし、および、論理レコード１７００とパリティレコード１５０１の書き込みという、ディスクアレイ適用以前の４倍の転送が発生する。

【００６２】ディスク装置１３０４側にパリティ作成機能を設けた場合の動作については、図２を用いて説明する。制御装置側１３０５では、ディスク装置１３０４から読み出した（２０６）論理レコード１７００（＝データレコード１５００）の更新前の値である更新前データ２０９と、処理装置１３００から受け取った（２０４）論理レコード１７００の更新後の値である論理レコード更新値２１０をパリティ中間値作成部２０１に入力し（２１２）、パリティレコードを作成するための中間値、すなわち、パリティ中間値２００を作成する。さらに、制御装置１３０５は、作成したパリティ中間値２００を、キャッシュ１３０８に格納する（２１３）。ただし、制御装置１３０５にキャッシュ１３０８が存在しない場合、制御装置バッファ１３１０に格納する。この後、制御装置１３０５は、パリティ中間値送信部２０２により、パリティ中間値２００をディスク装置１３０４に転送する（２１４）。

【００６３】ディスク装置１３０４側では、パリティ中間値入力部２０３により、制御装置１３０５から受け取ったパリティ中間値２００をバッファ１０２に格納する（２１４）。次に、パリティ読み出し部２０５により、パリティレコード１５０１の更新前の値を、記憶媒体１０１から読みだし、更新前パリティ２０７として、バッファ１０２に格納する（２１５）。さらに、パリティ中間値２００と更新前パリティ２０７をパリティ作成部ａ２０８に入力し（２１６）、パリティ更新値１０６を作成し、バッファ１０２に格納する（２１７）。この後、パリティレコードライト部１０５により、バッファ１０２に格納したパリティレコード１５０１の更新値を記憶媒体１０１に書き込む（１１２）。

【００６４】以上のような動作では、制御装置１３０５とディスク装置１３０４の間の転送処理は、データレコード１５００の読みだし、書き込み、および、パリティ中間値２００の転送となり、ディスクアレイ適用以前の３倍までに、転送量に抑えることができる。

【００６５】第２、第３の実施例では、ディスク装置１３０４に設けたパリティレコード作成機能と制御装置１３０５とディスク装置１３０４の間のブロードキャスト転送機能組み合わせて利用する。

【００６６】第２の実施例は、ディスク装置１３０４に設けたパリティレコード作成機能と制御装置１３０５とディスク装置１３０４の間のブロードキャスト転送機能とを、組み合わせて、分散配置のディスクアレイに適用した場合の実施例である。以下、図３を用いてその概要を説明する。

【００６７】この場合、制御装置１３０５は、論理レコードライト部３００により、パリティグループ１６００に属するｍ＋ｎ台のディスク装置１３０５に、処理装置１３００から受け取った論理レコード更新値２１０を分割せず、そのままの形で、ブロードキャストする（３０４）。この時、論理レコード更新値２１０を受け取るパリティグループ１６００内のディスク装置１３０４は、論理レコード更新値２１０の一部をデータレコード１５

００の更新値として記録すべきｍ台のデータディスク装置３０１と、更新対象となる論理レコード１７００に対応するパリティレコード１５０１を記録すべきｎ台のパリティディスク装置３０２に分類できる。ただし、データディスク装置３０１、および、パリティディスク装置３０２は、各ディスク装置１３０４に固定的な属性ではなく、転送対象となる論理レコード１７００ごとに決定される属性である。

【００６８】データディスク装置３０１の場合、まず、制御装置１３０５より受け取った論理レコード更新値２１０を、論理レコード入力部３０６により、バッファ１０２に格納する（３０４）。次に、データレコードライト部ａ３０３により、論理レコード更新値２１０の中から自装置で書き込むべきデータレコード１５００の更新値を取りだし、記憶媒体１０１に書き込む（３０７）。

【００６９】パリティディスク装置３０２の場合、論理レコード更新値２１０のバッファ１０２への格納（３０４）は、同様であるが、次にパリティ作成部ｂ３０５により、自装置で書き込むべきパリティ更新値１０６を、論理レコード１７００からを作成する（３０８）。この後、パリティレコードライト部１０５により、バッファ１０２に格納したパリティレコード１５０１の更新値を、記憶媒体１０１に書き込む（１１２）。

【００７０】パリティレコード作成機能とブロードキャスト転送機能を用いず、分割配置のディスクアレイを適用すると、制御装置１３０５とディスク装置１３０４の間で転送されるデータの増分は、パリティレコード１５０１の転送量となる。しかし、パリティレコード作成機能とブロードキャスト転送機能を用いた場合、制御装置１３０５が転送するのは論理レコード１７００のみであるため、ディスクアレイの適用に伴い生ずる制御装置１３０５とディスク装置１３０４の間のデータ転送路の転送オーバヘッドをなくすことが可能となる。

【００７１】図３に示した処理は、非分割配置のディスクアレイにおいても、パリティグループ１６００内のすべての論理レコード１７００（＝データレコード１５００）をまとめて更新する場合にも適用できる。というのは、図３に示した処理は、非分割配置のディスクアレイにおいて、制御装置１３０５が、パリティグループ１６００内のすべてのデータレコード１５００の更新値をブロードキャストするという処理に他ならないためである。

【００７２】第３の実施例は、ディスク装置１３０４に設けたパリティレコード作成機能と制御装置１３０５とディスク装置１３０４の間のブロードキャスト転送機能とを組み合わせて、非分散配置のディスクアレイに適用した場合の実施例である。以下、その概要を図４を用いて説明する。

【００７３】第３の実施例も、第２の実施例と同様に、論理レコードライト部３００により、パリティグループ１６００に属するｍ＋ｎ台のディスク装置１３０４に、処理装置１３００から受け取った論理レコード更新値２１０を、ブロードキャストする（３０４）。この場合、パリティグループ１６００内のディスク装置１３０４は、論理レコード１７００（＝データレコード１５００）をそのまま記録すべき１台のデータディスク装置３０１、何も処理をしなくてよいｍ－１台のディスク装置、および、論理レコード１７００に対応するパリティレコード１５０１を記録すべきｎ台のパリティディスク装置３０２に分類できる。第２の実施例と同様、データディスク装置３０１、パリティディスク装置３０２、および、何もしなくてもよいディスク装置１３０４は、ディスク装置１３０４に固定的な属性ではなく、論理レコード１７００対応に決定される属性である。また、何も処理をしなくてよいｍ－１台のディスク装置には、論理レコード更新値２１０を送らないようにしてもよい。なお、何も処理をしなくてよいｍ－１台のディスク装置は、特に説明すべき点がないため図４には示さなかった。

【００７４】パリティディスク装置３０２は、まず、第２の実施例と同様、論理レコード入力部３０６により、論理レコード更新値２１０をバッファ１０２に格納する（３０４）。次に、データレコード受信部４０２により、論理レコード１７００（＝データレコード１５００）の更新前の値である更新前データ２０９を、論理レコード１７００（＝データレコード１５００）を記録しているデータディスク装置３０１から、受け取り、バッファ１０２に格納する（４０４）。さらに、パリティ読み出し部２０５により、更新前データ２０９を記憶媒体１０１から読みだす（２１５）。以上のようにして得た、論理レコード更新値２１０、更新前データ２０９及び更新前パリティ２０７をパリティ作成部ｃ４００に入力し、パリティ更新値１０６を作成する（４０５）。この後、パリティレコードライト部１０５により、バッファ１０２に格納したパリティレコード１５０１の更新値１０６を記憶媒体１０１に書き込む（１１２）。

【００７５】データディスク装置３０１の場合も、まず、第２の実施例と同様、論理レコード入力部３０６により、論理レコード更新値２１０をバッファ１０２に格納する。次に、ｎ台のパリティディスク装置３０２に、更新前データ２０９を送信する。この後、データレコードライト部ｂ４０１により、制御装置１３０５から受け取った論理レコード更新値２１０をデータレコード１５００の更新値として、記憶媒体１０１に書き込む。

【００７６】以上の動作では、制御装置１３０５とディスク装置１３０４の間のデータ転送路を用いて転送されるデータは、論理レコード１７００の更新前後の値である。したがって、この場合、制御装置１３０５とディスク装置１３０４の間のデータ転送量を、ディスクアレイの適用以前の２倍までに抑えることができる。

【００７７】以下、各実施例の詳細を説明する。まず、第１の実施例の詳細について説明する。

【００７８】第１の実施例は、ディスク装置１３０４にパリティレコード作成機能を設け、非分散配置をとったディスクアレイに関する。第１の実施例において、制御装置１３０５とディスク装置１３０４の間のデータ転送路の転送量の削減を可能にするのは、制御装置１３０５が、処理装置１３００から受け取った論理レコード１７００（＝データレコード１５００）の更新値と、論理レコード１７００の更新前の値、すなわち、更新前データ２０９から、パリティ中間値２００を作成し、ディスク装置１３０４に転送する点である。

【００７９】以下、以上の機能を実行する図２に示したパリティ中間値作成部２０１とパリティ中間値送信部２０２の処理フローについて説明する。以上の処理は、処理装置１３００から受け取った論理レコード１７００を更新するためのライト要求と同期して実行してもよいし、あるいは、ライト要求の完了報告を返した後、非同期に実行してもよい。ただし、同期、非同期にかかわらず、少なくとも処理装置１３００から受け取った論理レコード１７００を、制御装置メモリ１３１０、あるいは、キャッシュ１３０８に格納した後、実行を、開始する必要がある。

【００８０】図２に示したパリティ中間値作成部２０１の処理フローを図５に示す。図５は、開始点ａと開始点ｂの２つの開始点をもつ。制御装置１３０５は、まず、開始点ａから実行を開始する。ステップ５００で、ディレクトリ１３０９を参照して、キャッシュ１３０８に、更新前データ２０９が格納されているかをチェックする。この時、制御装置１３０５が、キャッシュ１３０８をもたず、制御装置メモリ１３１０をもつ場合、直ちに、ステップ５０１の実行にはいる。存在する場合、ステップ５０３へジャンプする。存在しない場合、ステップ５０１で、ライト対象となる論理レコード１７００に対応するデータレコード１５００を格納したディスク装置１３０４に、位置付け要求を発行し、一度処理を終了する。

【００８１】開始点ｂは、ディスク装置１３０４の位置付け処理が完了した時、実行を開始するポイントである。ステップ５０２では、制御装置１３０４は、ディスク装置１３０４から、更新前データ２０９を、キャッシュ１３０８、あるいは、制御装置メモリ１３１０にロードする。キャッシュ１３０８にロードした場合、ディレクトリ１３０９の内容を更新する。

【００８２】ステップ５０３では、制御装置１３０５は、キャッシュ１３０８、あるいは、制御装置メモリ１３１０に格納した論理レコード１７００の更新値と更新前データ２０９より、パリティ中間値２００を作成し、キャッシュ１３０８、あるいは、制御装置メモリ１３１０に格納する。

【００８３】図３０は、パリティ中間値送信部２０２の処理フローである。ステップ３０００では、制御装置１３０５は、ｎ台存在するパリティレコードを更新すべきディスク装置の内、まだパリティ中間値２００を送っていない１台に、パリティ中間値２００を送信する。ステップ３００１では、ｎ台すべてに送ったかどうかをチェックする。そうであれば、処理を終了し、そうでなければステップ３０００に制御を戻す。

【００８４】次に、図２に示したディスク装置１３０５側の各処理部の処理フローを説明する。

【００８５】まず、図６を用いて、パリティ中間値入力部２０３の処理フローを説明する。ステップ６００で、プロセッサ１００は、制御装置１３０５から受け取ったパリティ中間値２００をバッファ１０２に格納する。この時、バッファ管理情報１０７を、パリティ中間値２００のバッファ１０２への格納に対応して、更新する。次に、図７を用いて、パリティ読み出し部２０５の処理フローを説明する。まず、プロセッサ１００は、ステップ７００で、バッファ管理情報１０７を参照し、更新前パリティ２０７が、バッファ１０２に存在するかを、チェックする。存在する場合、処理を終了し、存在しない場合、ステップ７０１の実行に入る。バッファ１０２の容量が小さく、更新前パリティ２０７の存在が期待できない場合、直ちにステップ７０１から実行を開始してもよい。ステップ７０１では、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ７０２で、更新前パリティ２０７をバッファ１０２に格納する。この時、バッファ管理情報１０７を、更新前パリティ２０７のバッファ１０２への格納に対応して、更新する。この後、処理を終了する。

【００８６】図８は、パリティ作成部ａ２０８の処理フローである。ステップ８００で、プロセッサ１００は、パリティ中間値２００と更新前パリティ２０７からパリティ更新値１０６を作成し、バッファ１０２に格納する。この時、図１に示すバッファ管理情報１０７を、パリティ更新値１０６のバッファ１０２への格納に対応して、更新する。この後、処理を終了する。

【００８７】図９は、パリティレコードライト部１０５の処理フローである。まず、ステップ８００で、プロセッサ１００は、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ８０１で、バッファ１０２に格納したパリティレコード１５０１の更新値を記憶媒体１０１に書き込む。この後、処理を終了する。

【００８８】次に、第２の実施例の詳細について説明する。

【００８９】第２の実施例は、ディスク装置１３０４に設けたパリティレコード作成機能と制御装置１３０５とディスク装置１３０４の間のブロードキャスト転送機能

とを組み合わせて、分散配置のディスクアレイに適用した場合の実施例である。以下、図３に示した制御装置１３０４とデータディスク装置３０１とパリティディスク装置３０２内の各処理部の内容を処理フローを用いて説明する。

【００９０】図１０は、制御装置１３０５内の図３に示す論理レコードライト部３００の処理フローである。本処理は、処理装置１３００から受け取った論理レコード１７００を更新するためのライト要求と同期して実行してもよいし、あるいは、ライト要求の完了報告を返した後、非同期に実行してもよい。この場合、制御装置１３０５は、ステップ１０００で、パリティグループ１６００に属するｍ＋ｎ台のディスク装置１３０５に、処理装置１３００から受け取った論理レコード１７００を分割せず、そのままの形で、ブロードキャストする。

【００９１】次に、データディスク装置３０１の各処理部について説明する。

【００９２】図１１は、論理レコード入力部３０６の処理フローである。プロセッサ１００は、ステップ１１００で、制御装置２０５より受け取った論理レコード１７００を、バッファ１０２に格納する。この時、バッファ管理情報１０７を、論理レコード１７００のバッファ１０２への格納に対応して、更新する。この後、処理を終了する。

【００９３】図１２は、データレコードライト部ａ３０３の処理フローである。まず、ステップ１２００で、プロセッサ１００は、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ１２０１で、論理レコード１７００の中から自装置で書き込むべき部分をデータレコード１５００として取りだし、記憶媒体１０１に書き込む。この後、処理を終了する。

【００９４】最後に、パリティディスク装置３０２の各処理部について説明する。

【００９５】論理レコード入力部３０６の処理フローは、データディスク装置３０１の場合と共通であるため、説明を省略する。

【００９６】図１９は、図３に示すパリティ作成部ｂ３０５の処理フローである。ステップ１９００で、プロセッサ１００は、自装置で書き込むべきパリティ更新値１０６を、論理レコード１７００からを作成し、バッファ１０２に格納する。この時、バッファ管理情報１０７を、パリティ更新値１０６のバッファ１０２への格納に対応して、更新する。この後、処理を終了する。

【００９７】パリティレコードライト部１０５の処理フローは、第１の実施例と共通であるため、説明を省略する。

【００９８】最後に、第３の実施例の詳細について説明する。

【００９９】第２の実施例は、ディスク装置１３０４に設けたパリティレコード作成機能と制御装置１３０５とディスク装置１３０４の間のブロードキャスト転送機能と、分散配置組み合わせて利用した場合の実施例である。以下、図４に示した制御装置１３０４とデータディスク装置３０１とパリティディスク装置３０２内の各処理部の内容を処理フローを用いて説明する。ただし、制御装置１３０５内の論理レコードライト部３００の処理フローは、第２の実施例と共通であるため、説明を省略する。

【０１００】まず、パリティディスク装置３０２の各処理部について説明する。ただし、論理レコード入力部３０６の処理フローは、第２の実施例と共通であるため、説明を省略する。

【０１０１】図２０と図２１は、データレコード受信部ａ４０２の処理フローである。プロセッサ１００は、図２０の処理フローから実行を開始する。ステップ２０００で、バッファ管理情報１０７を参照して、データレコード１５００の更新前の値である更新前データ２０９が、バッファ１０７内に存在するかをチェックする。存在する場合は、ステップ２００１で、更新前データ２０９を格納したデータディスク装置３０１に、更新前データ２０９の送信要求を発行する。この後、一度処理を終了する。バッファ１０２の容量が少なく、更新前データ２０９の存在が、期待できない場合、直ちに、ステップ２００１を実行してもよい。存在する場合、プロセッサ１００は、ステップ２００２で、更新前データ２０９を格納したデータディスク装置３０１に、更新前データ２０９の送信が必要ないことを通知する。この後、処理を終了する。

【０１０２】図２１の処理フローは、更新前データ２０９を格納したデータディスク装置３０１から、更新前データ２０９を送られてきた時、実行を開始する処理を表した処理フローである。ステップ２１００で、プロセッサ１００は、更新前データ２０９を、バッファ１０２に格納する。この時、バッファ管理情報１０７を、更新前データ２０９の格納に対応して、更新する。この後、処理を終了する。

【０１０３】パリティ入力部２０５の処理フローは、第１の実施例と共通であるため、説明を省略する。

【０１０４】図２２は、パリティ作成部ｃ４００の処理フローである。プロセッサ１００は、ステップ２２００で、論理レコード更新値２１０、更新前データ２０９および更新前パリティ２０７より、パリティ更新値１０６を作成し、バッファ１０２に格納する。この時、バッファ管理情報１０７を、パリティ更新値１０６のバッファ１０２への格納に対応して、更新する。この後、処理を終了する。

【０１０５】パリティレコードライト部１０５の処理フローは、第１の実施例と共通であるため、説明を省略する。

【０１０６】次に、データディスク装置３０１の各処理部について説明する。ただし、論理レコード入力部３０６の処理フローは、第２の実施例と共通であるため、説明を省略する。

【０１０７】図２３は、データレコード送信部ａ４０３の処理フローである。ステップ２３００で、プロセッサ１００は、ｎ台のパリティディスク装置３０２から、更新前データ２０９の送信要求がくるまで待つ。ｎ台のパリティディスク装置３０２からの要求が到着すると、ステップ２３０１で、プロセッサ１００は、ｎ台すべてのパリティディスク装置３０２が、更新前データ２０９の送信を必要としないかをチェックする。そうなら処理を終了する。少なくとも１台のパリティディスク装置３０２が、更新前データ２０９の送信を必要とするなら、ステップ２３０２で、プロセッサ１００は、バッファ管理情報１０７を参照して、更新前データ２０９が、バッファ１０２内に存在するかをチェックする。存在する場合は、ステップ２３０５へジャンプする。バッファ１０２の容量が少なく、更新前データ２０９の存在が、期待できない場合、直ちに、ステップ２３０２をスキップして、直接ステップ２３０３を実行してもよい。存在しない場合、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ２３０４で、更新前データ２０９をバッファ１０２に格納する。この時、バッファ管理情報１０７を、更新前データ２０９のバッファ１０２への格納に対応して、更新する。

【０１０８】ステップ２３０５では、プロセッサ１００は、ｎ台のパリティディスク装置３０２のうち更新前データ２０９を必要とするパリティディスク装置３０２に、更新前データ２０９を、ブロードキャストする。この後、処理を終了する。ここで、ステップ２３０５で、すべてのパリティディスク装置３０２に、更新前データ２０９の送信し、更新前データ２０９が存在するパリティディスク装置３０２は、送られてきた更新前データ２０９を捨てるようにしてもよい。

【０１０９】図２４は、データレコードライト部ｂ４０１の処理フローである。まず、ステップ２４００で、プロセッサ１００は、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ２４０１で、により、制御装置１３０５から受け取った論理レコード１７００をデータレコード１５００として、記憶媒体１０１に、書き込む。この後、処理を終了する。

【０１１０】何もしなくてよいｍ－１台のディスク装置１３０４については、特に説明すべきことはないため、説明は行わない。

【０１１１】第３の実施例では、制御装置１３０５とディスク装置１３０４の間のデータ転送路を用いて転送されるデータは、最大限、論理レコード１７００の更新値と更新前データ２０９である。しかし、ステップ２０００から明らかのように、更新前データ２０９が、パリティディスク装置３０２のバッファ１０１内に存在すれば、更新前データ２０９は、制御装置１３０５とディスク装置１３０４の間のデータ転送路を用いて転送する必要はない。通常、論理レコード１７００が、処理装置１３００からライト対象になる直前に、処理装置１３００からリードされることが多い。したがって、データディスク装置３０１が、処理装置１３００からリード要求をうけた論理レコード１７００（＝データレコード１５００）を制御装置１３０５に送るとき、同時に、パリティディスク装置３０２にもブロードキャストしておくことが有効となる。というのは、論理レコード１７００が、ライト対象となった時、更新前データ２０９が、パリティディスク装置３０２のバッファ１０１内に存在する確率を高めることができるためである。あるいは、無条件にブロードキャストを行うのではなく、処理装置１３００から、制御装置１３０５経由で、当該リード要求でリードされた論理レコード１７００は、この後、ライトされる確率が高いという情報を受け、この時だけ、論理レコード１７００を、パリティディスク装置３０２にブロードキャストするようにしてもよい。

【０１１２】図２５は、データディスク装置３０１が、論理レコード１７００（＝データレコード１５００）を制御装置１３０５に送るとき、同時に、パリティディスク装置３０２にもブロードキャストする動作を表している。

【０１１３】データディスク装置３０１内のデータレコード送信部ｂ２５００は、制御装置１３０５から、データレコード１５００の送信要求を受けたとき、論理レコード１７００（＝データレコード１５００）を、制御装置１３０５とパリティディスク装置３０２にもブロードキャストする（２５０３）。ただし、データレコード送信部ｂ２５００が、上記のブロードキャストを行うのは、論理レコード１７００の送信要求を受けた時、常に実行するのではなく、その論理レコード１７００が後でライトされる確率が高いという指示を、制御装置１３０５から通知された時だけでもよい。

【０１１４】制御装置１３０５内の、論理レコードリード部２５０１は、データディスク装置３０１から送られてきた論理レコード１７００（＝データレコード１５００）を、処理装置１７００に送る。

【０１１５】パリティディスク装置３０２内の、データレコード入力部ａ２５０２では、送られてきた論理レコード１７００（＝データレコード１５００）を、更新前データ２０９としてバッファ１０２に格納する。

【０１１６】以下、各処理部の詳細を処理フローを用いて詳細に説明する。

【０１１７】図２６は、データディスク装置３０１内のデータレコード送信部ｂ２５００の処理フローである。

ステップ２６００で、プロセッサ１００は、バッファ管理情報１０７を参照して、更新前データ２０９が、バッファ１０２内に存在するかをチェックする。存在する場合は、ステップ２３０５へジャンプする。バッファ１０２の容量が少なく、更新前データ２０９の存在が、期待できない場合、直ちに、直接ステップ２６０１を実行してもよい。存在しない場合、ステップ２６０１で、記憶媒体１０１の位置付け処理が完了するのをまつ。位置付け処理が完了した時、プロセッサ１００は、ステップ２６０２で、更新前データ２０９をバッファ１０２に格納する。この時、バッファ管理情報１０７を、更新前データ２０９のバッファ１０２への格納に対応して、更新する。

【０１１８】ステップ２６０３では、プロセッサ１００は、制御装置１３０５とｎ台のパリティディスク装置３０２に、更新前データ２０９を、ブロードキャストする。この後、処理を終了する。

【０１１９】図２７は、制御装置１３０５内の、論理レコードリード部２５０１の処理フローである。まず、ステップ２７００で、プロセッサ１００は、データディスク装置３０１から送られてきた論理レコード１７００（＝データレコード１５００）を、処理装置１７００に送る。この時、送られてきた論理レコード１７００をキャッシュ１３０８に格納してもよい。キャッシュ１３０８に格納した場合、これに対応し、ディレクトリ１３０９を更新する。

【０１２０】図２９は、パリティディスク装置３０２内の、データレコード受信部ｂ２５０２の処理フローである。ステップ２９００で、プロセッサ１００は、受け取った論理レコード１７００（＝データレコード１５００）を、更新前データ２０９として、バッファ１０２に格納する。この時、バッファ管理情報１０７を、更新前データ２０９の格納に対応して、更新する。この後、処理を終了する。

【０１２１】

【発明の効果】ディスクアレイを用いた場合、処理装置からライト処理を受け付けた場合、論理レコードの内容の変化に伴い、パリティレコードの内容も書き換える必要が生ずる。したがって、ライト対象となった論理レコードの更新値以外に、パリティレコードの更新値を作成するのに必要な情報を揃えるための転送処理と、パリティレコードの書き込みを行う転送処理が、制御装置とディスク装置の間のデータ転送路に発生することになる。本発明では、ディスク装置にパリティレコード作成機能を設けることにより、パリティレコードの操作に起因する、制御装置とディスク装置の間のデータ転送路の転送オーバヘッドを抑えることを可能にする。

【図面の簡単な説明】

【図１】本発明の概要

【図２】第１の実施例の概要

【図３】第２の実施例の概要

【図４】第３の実施例の概要

【図５】パリティ中間値送信部２０２の処理フロー

【図６】パリティ中間値入力部２０３の処理フロー

【図７】パリティ読み出し部２０５の処理フロー

【図８】パリティ作成部ａ２０８の処理フロー

【図９】パリティレコードライト部１０５の処理フロー

【図１０】論理レコードライト部３００の処理フロー

【図１１】論理レコード入力部３０６の処理フロー

【図１２】図１２は、データレコードライト部ａ３０３の処理フロー

【図１３】本発明の対象となる計算機システムの構成

【図１４】本発明の対象となる別の計算機システムの構成

【図１５】ディスク装置１３０４に記録されている物理レコード１５０２の構成

【図１６】ディスクアレイにおけるパリティグループ１６００の構成

【図１７】分散配置のディスクアレイのレコード配置

【図１８】非分散配置のディスクアレイのレコード配置

【図１９】パリティ作成部ｂ３０５の処理フロー

【図２０】データレコード入力部ａ４００の処理フロー

【図２１】データレコード入力部ａ４００の処理フロー

【図２２】パリティ作成部ｃ４０２の処理フロー

【図２３】データレコード送信部ａ４０２の処理フロー

【図２４】データレコードライト部ｂ４０３の処理フロー

【図２５】データディスク装置３０１が、論理レコード１７００（＝データレコード１５００）を制御装置１３０５に送るとき、同時に、パリティディスク装置３０２にもブロードキャストする動作を表す

【図２６】データレコード送信部ｂ２５００の処理フロー

【図２７】論理レコードリード部２５０１の処理フロー

【図２８】制御装置内に２つ以上のディレクタ１３０７を含む構成

【図２９】データレコード入力部ｂ２５０２の処理フロー

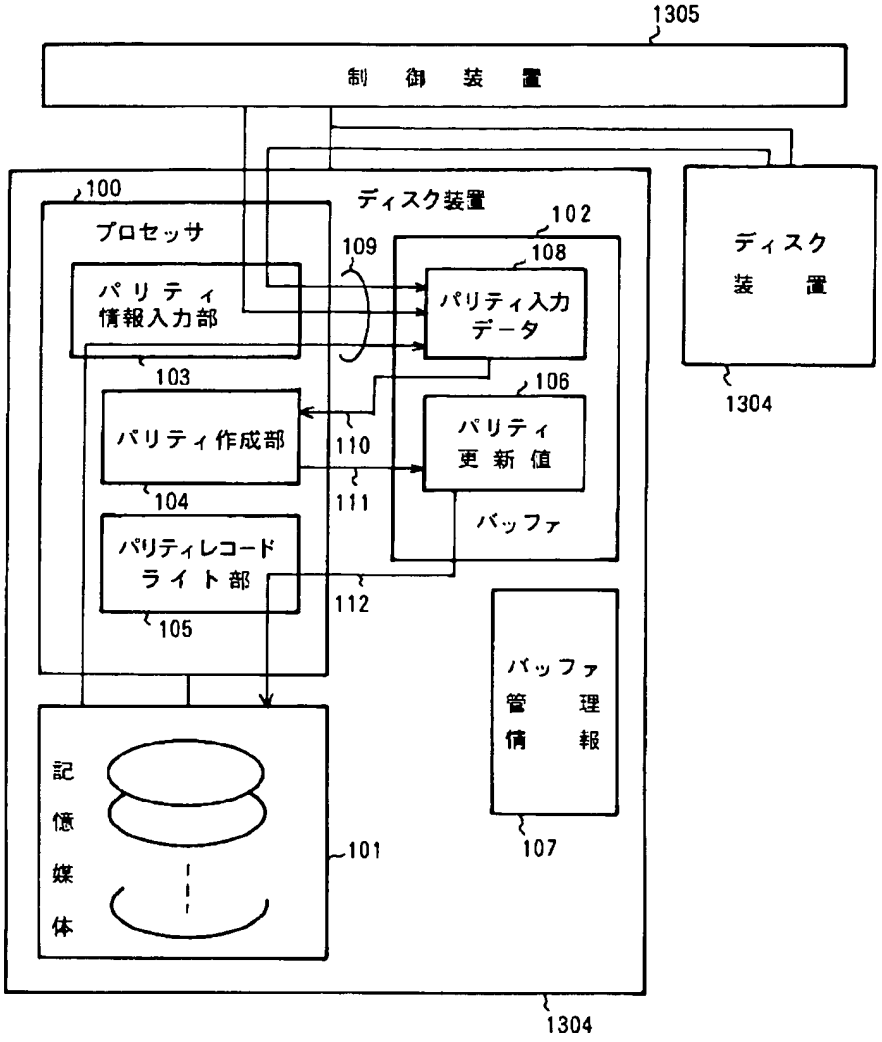
【図３０】パリティ中間値送信部２０２の処理フロー

【符号の説明】

１０３…パリティ情報入力部、１０４…パリティ作成部、１０５…パリティライト部、１０２…バッファ、１０７…バッファ管理情報。

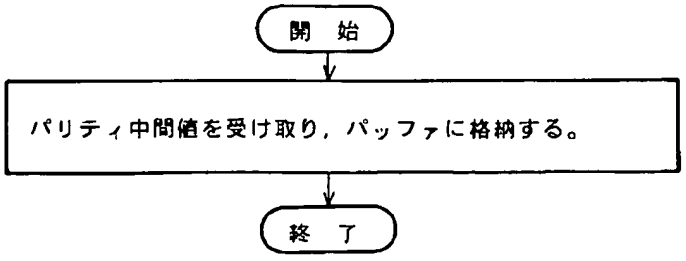
【図１】

図　1


1305
制　御　装　置
100
プロセッサ
ディスク装置
102
109
108
パリティ
情報入力部
パリティ入力
データ
ディスク
装　置
103
106
パリティ作成部
110
パリティ
更新値
104
111
バッファ
パリティレコード
ライト部
112
105
バッファ
管　理
情　報
記
憶
媒
体
101
107
1304
1304

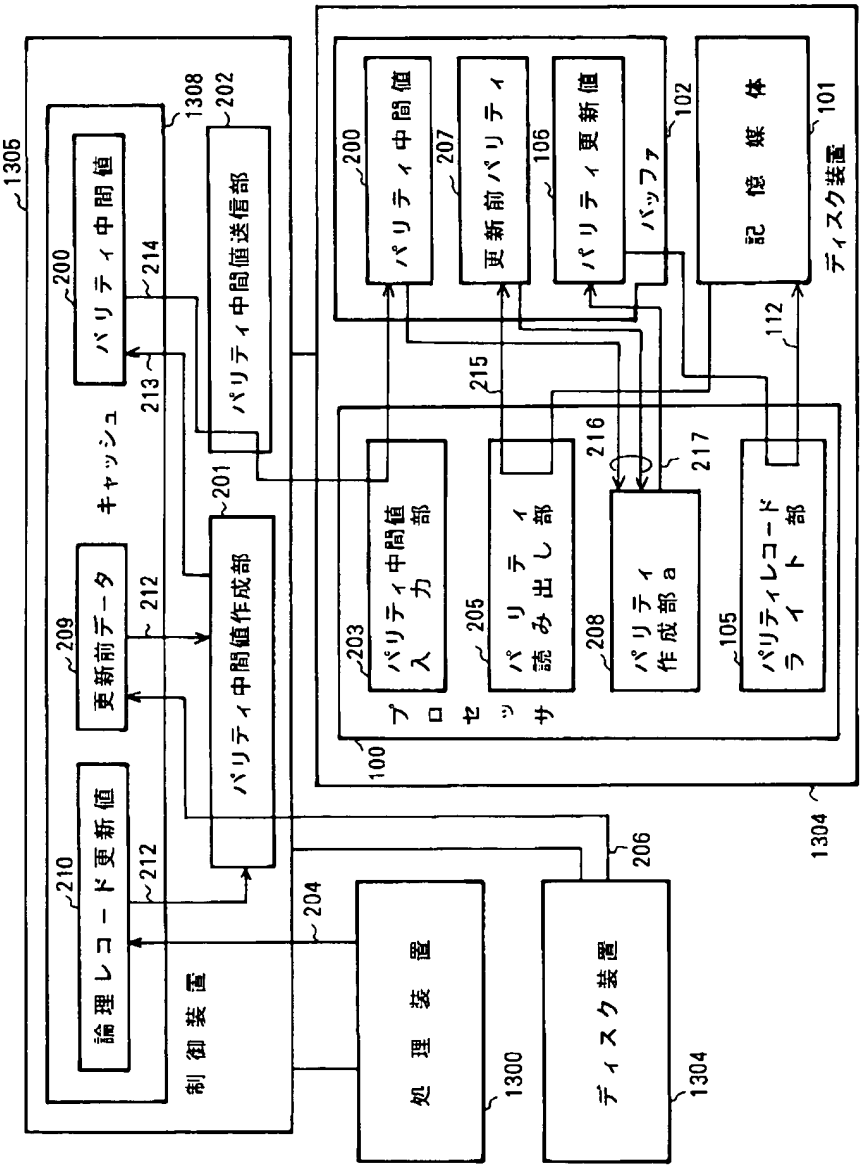
【図６】

図 6


開　始
パリティ中間値を受け取り，バッファに格納する。
終　了

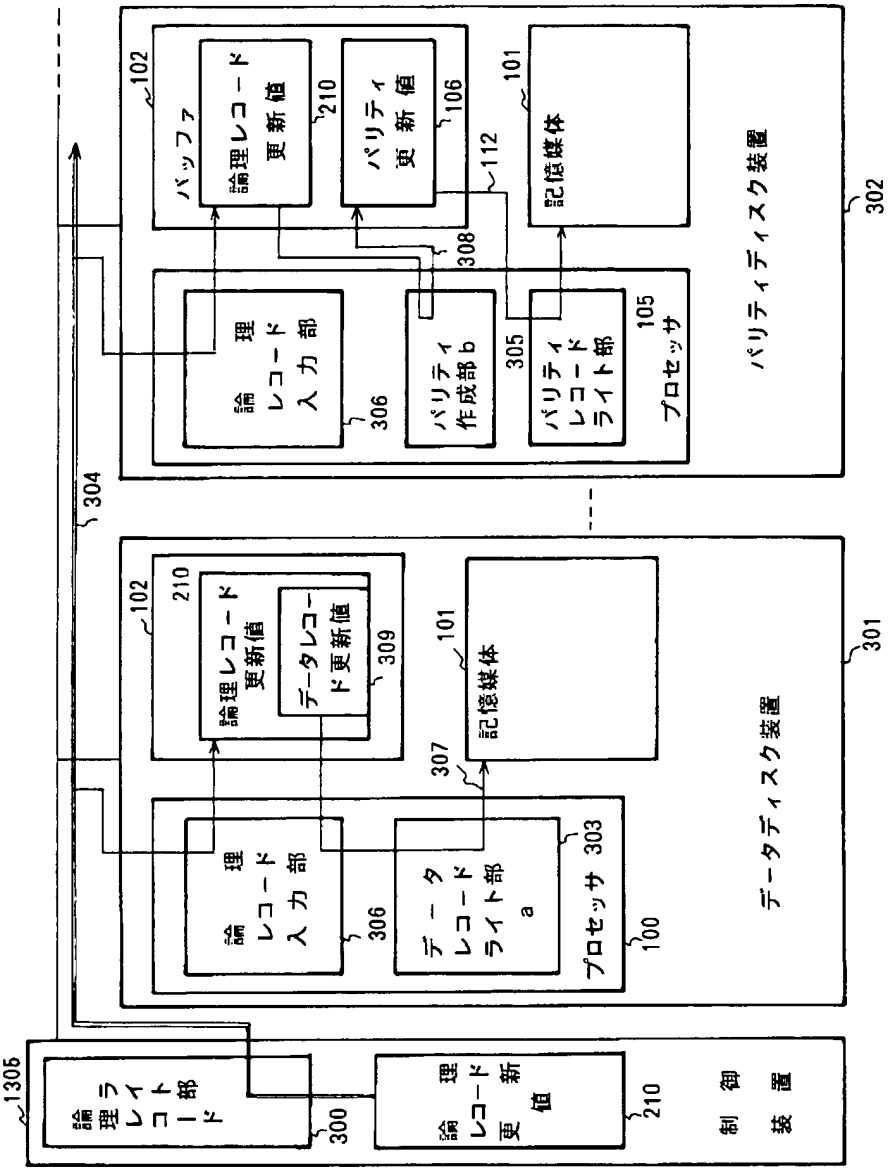
【図２】

図　２


1305
制御装置
論理レコード更新値
210
212
更新前データ
209
212
パリティ中間値作成部
201
204
キャッシュ
1308
パリティ中間値
200
214
213
パリティ中間値送信部
202
処理装置
1300
ディスク装置
1304
206
プロセッサ
100
パリティ中間値入力部
203
パリティ読み出し部
205
パリティ作成部a
208
パリティレコードライト部
105
215
216
217
パリティ中間値
200
207
更新前パリティ
106
パリティ更新値
バッファ
102
記憶媒体
ディスク装置
101
112
1304

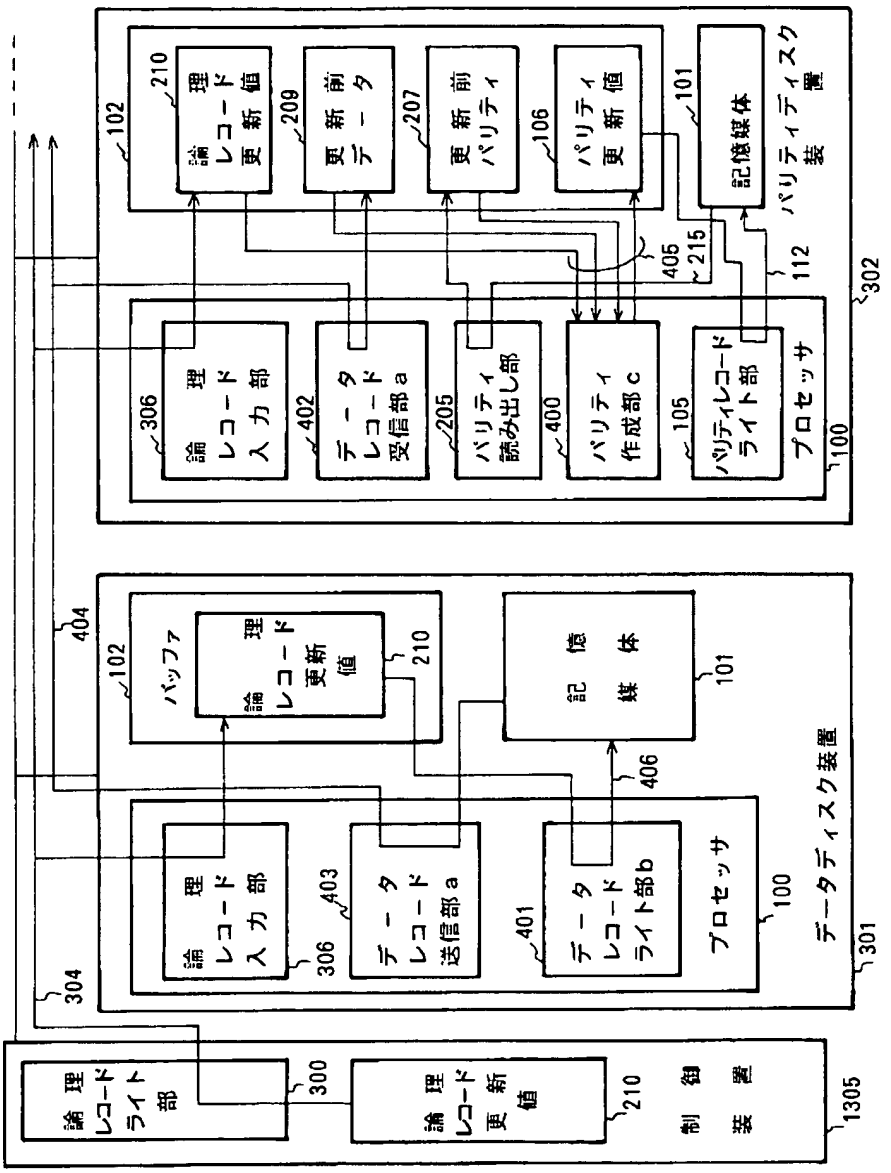
【図３】

図　３


制御装置
論理ライトレコード
300
論理レコード更新値
210
1306
データディスク装置
301
論理レコード入力部
306
データレコードライト部 a
307
プロセッサ 303
100
記憶媒体
101
論理レコード更新値
210
データレコード更新値
309
102
パリティディスク装置
302
論理レコード入力部
306
パリティ作成部 b
308
パリティレコードライト部
305
プロセッサ
105
記憶媒体
101
パリティ更新値
106
論理レコード更新値
210
バッファ
102
112
304

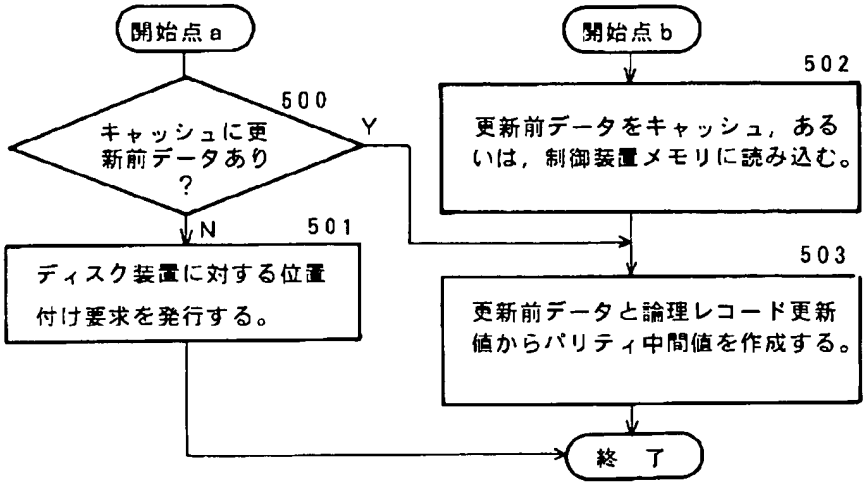
【図４】

図　４


制御装置
論理レコードライト部
論理レコード更新値
データディスク装置
論理レコード入力部
データレコード送信部a
データレコードライト部b
プロセッサ
バッファ
記憶媒体
パリティディスク装置
パリティレコードライト部
パリティ作成部c
パリティ読み出し部
データレコード受信部a
更新前データ
更新前パリティ
パリティ更新値
記憶媒体
300
301
302
304
306
403
401
404
406
102
210
101
105
400
205
402
209
207
106
215
112
100

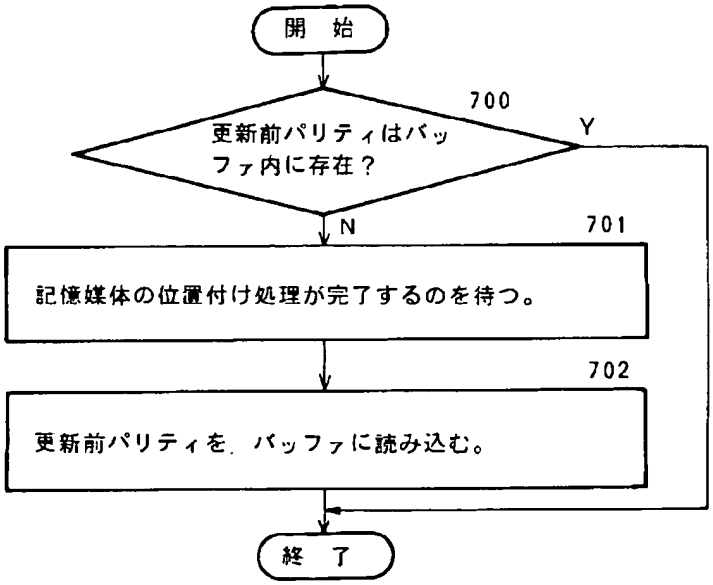
【図5】

図5


開始点a
開始点b
500
キャッシュに更新前データあり？
Y
N
501
ディスク装置に対する位置付け要求を発行する。
502
更新前データをキャッシュ，あるいは，制御装置メモリに読み込む。
503
更新前データと論理レコード更新値からパリティ中間値を作成する。
終　了

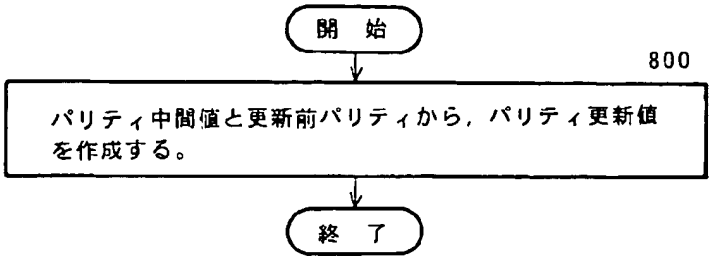
【図7】

図7


開　始
700
更新前パリティはバッファ内に存在？
Y
N
701
記憶媒体の位置付け処理が完了するのを待つ。
702
更新前パリティを，バッファに読み込む。
終　了

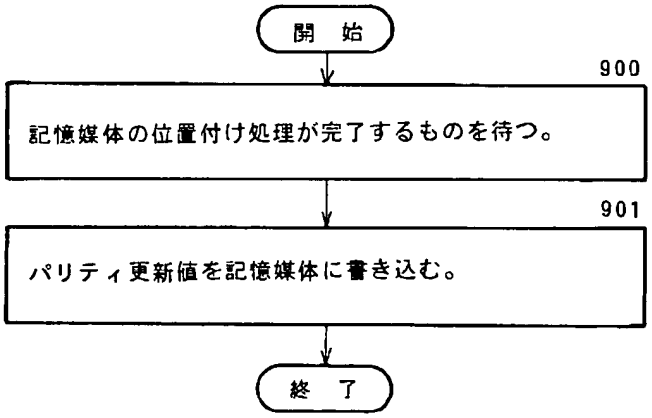
【図８】

図８



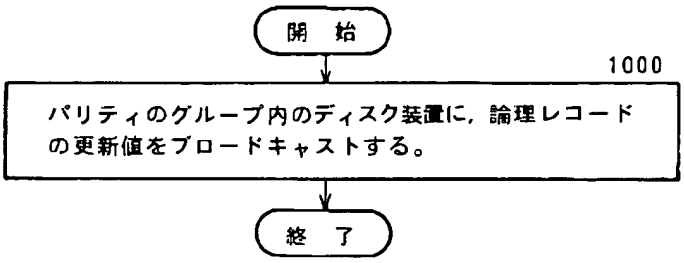
【図９】

図９



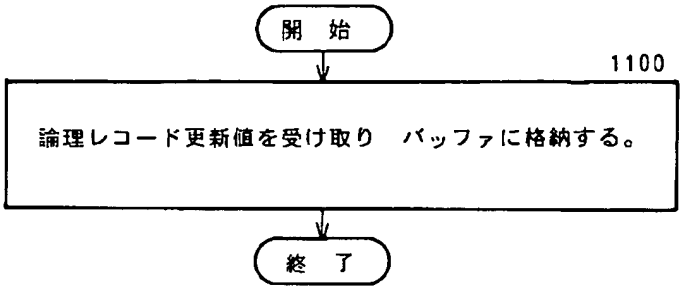
【図１０】

図１０



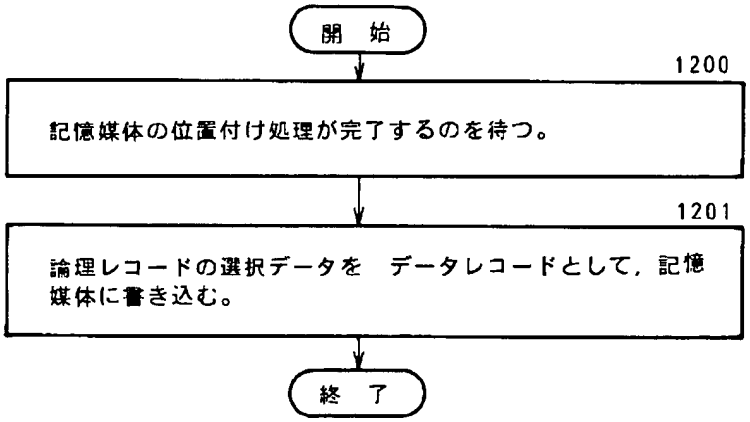
【図１１】

図１１



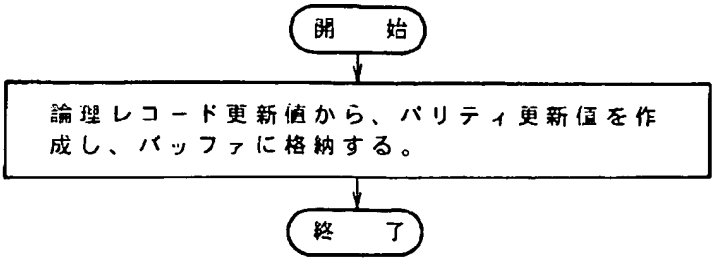
【図１２】

図１２



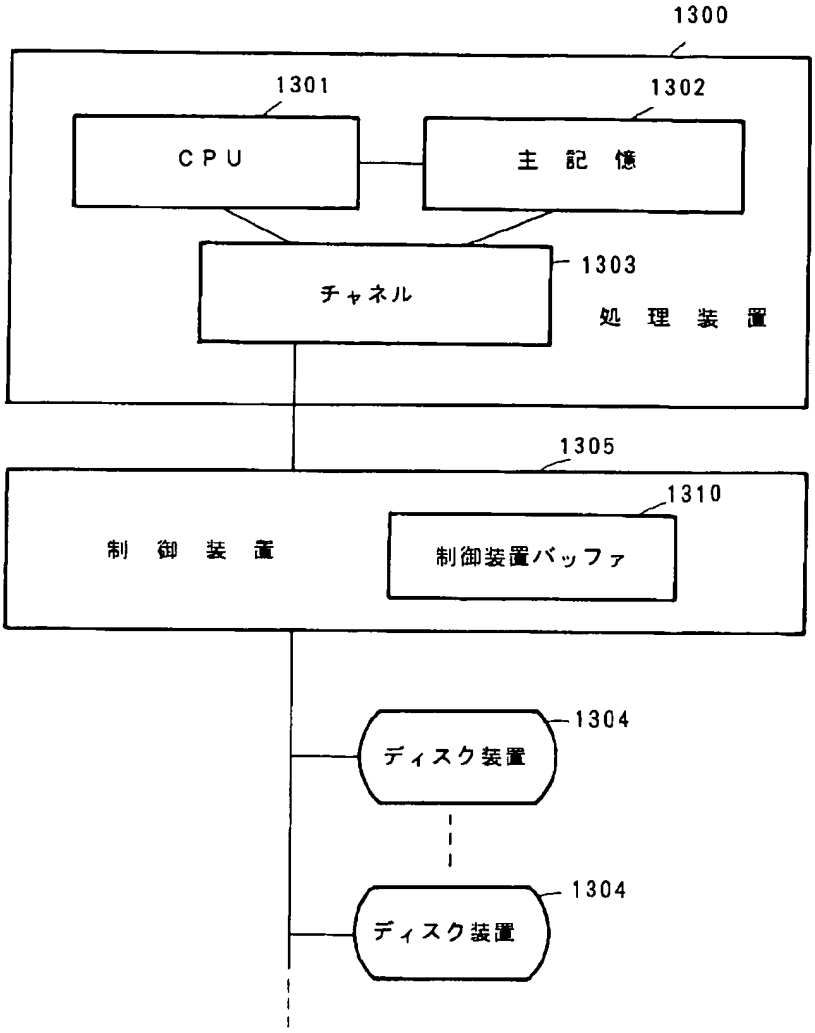
【図１９】

図１９



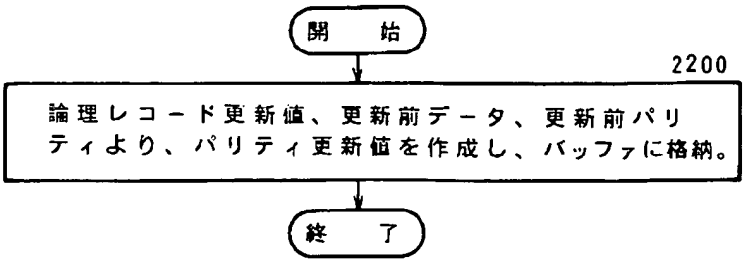
【図１３】

図１３



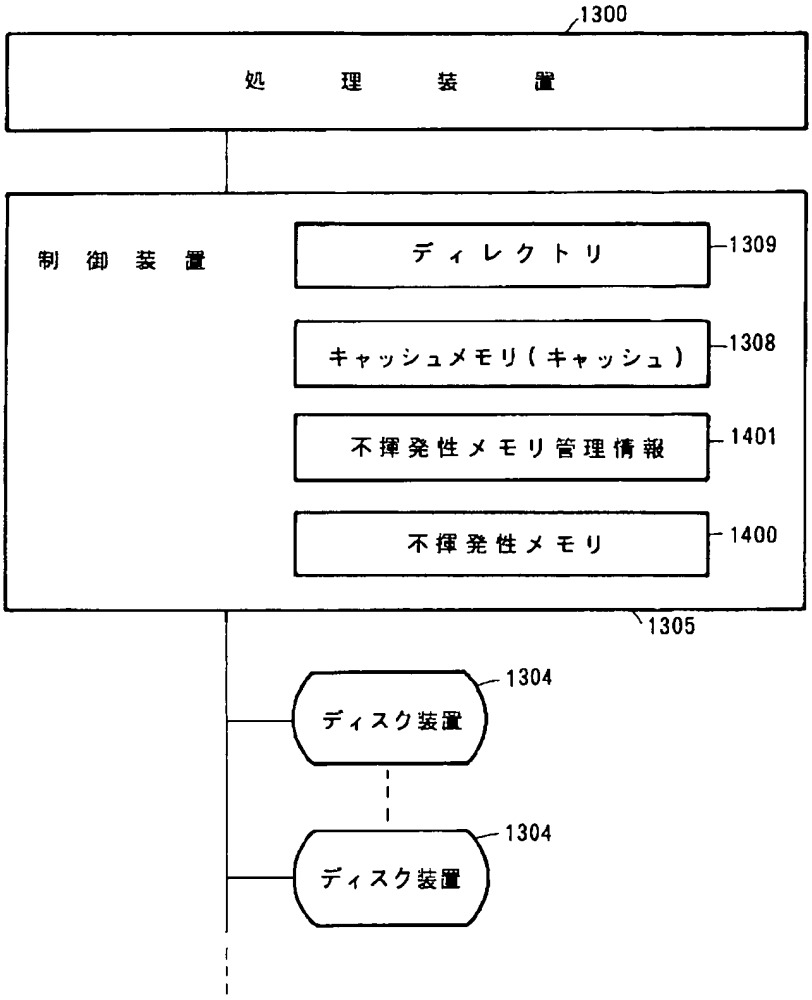
【図２２】

図２２



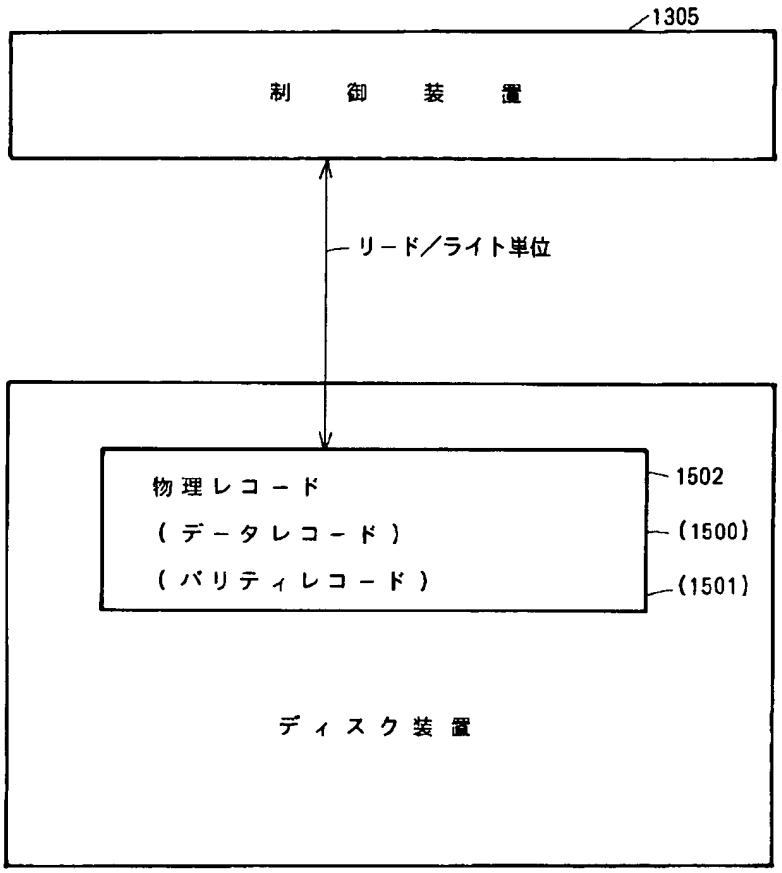
【図１４】

図１４


1300
処　理　装　置
制　御　装　置
ディレクトリ
1309
キャッシュメモリ（キャッシュ）
1308
不揮発性メモリ管理情報
1401
不揮発性メモリ
1400
1305
ディスク装置
1304
ディスク装置
1304

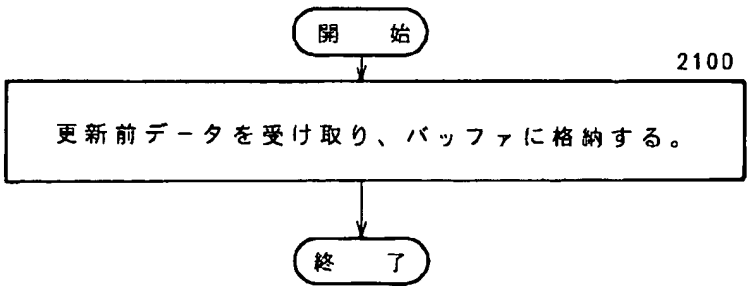
【図１５】

図１５



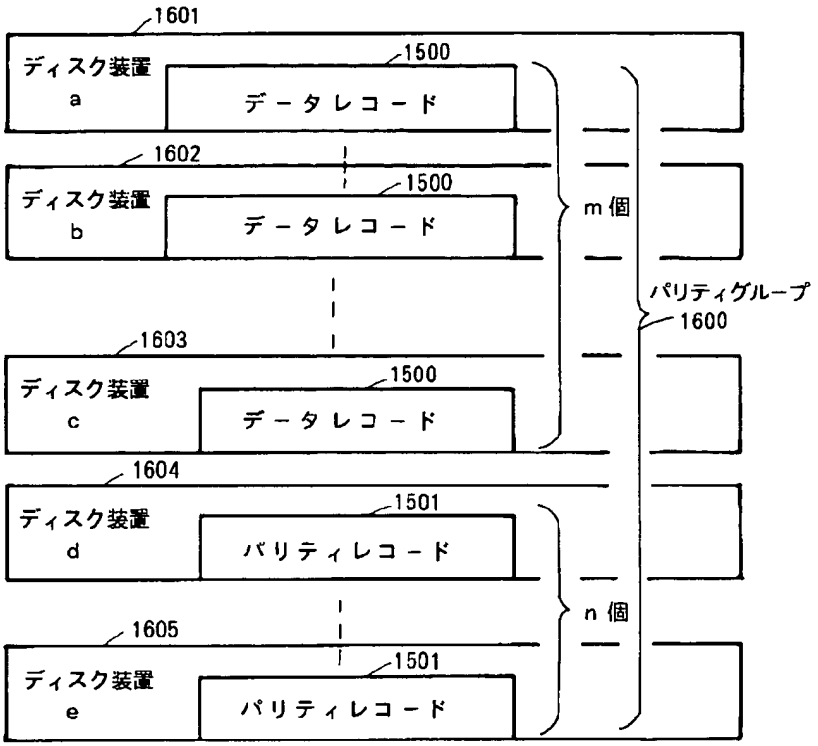
【図２１】

図２１



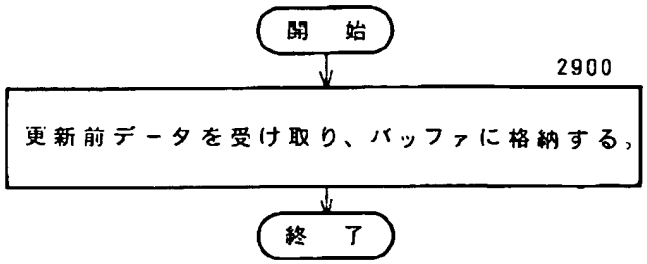
【図１６】

図１６



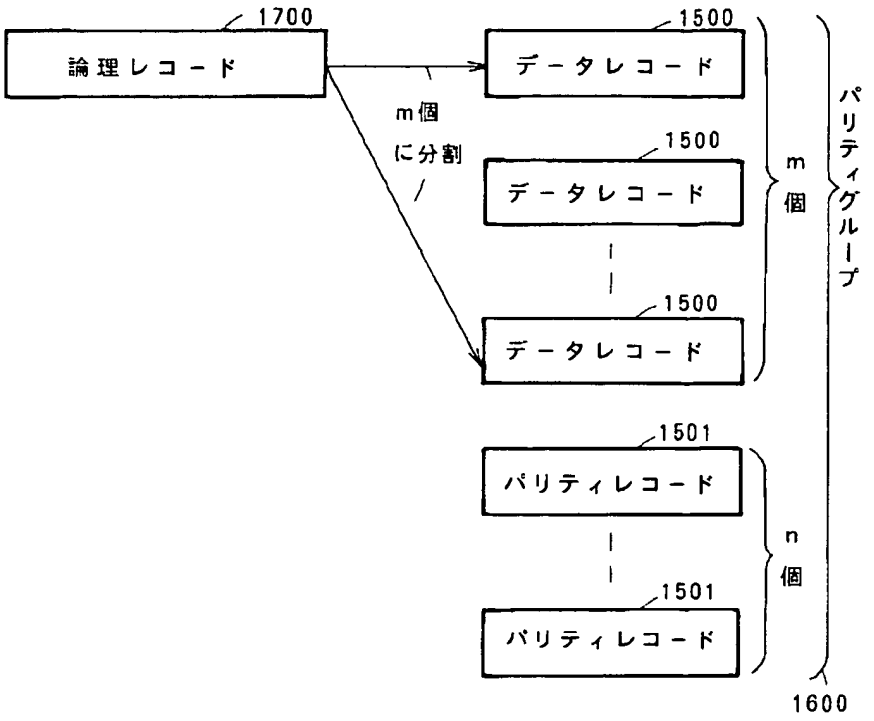
【図２９】

図２９



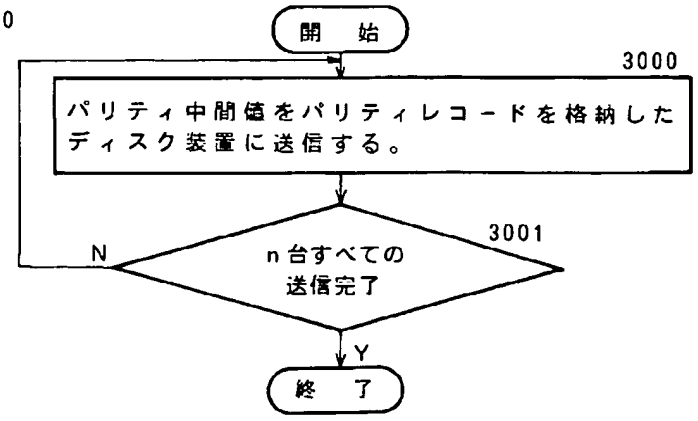
【図１７】

図１７



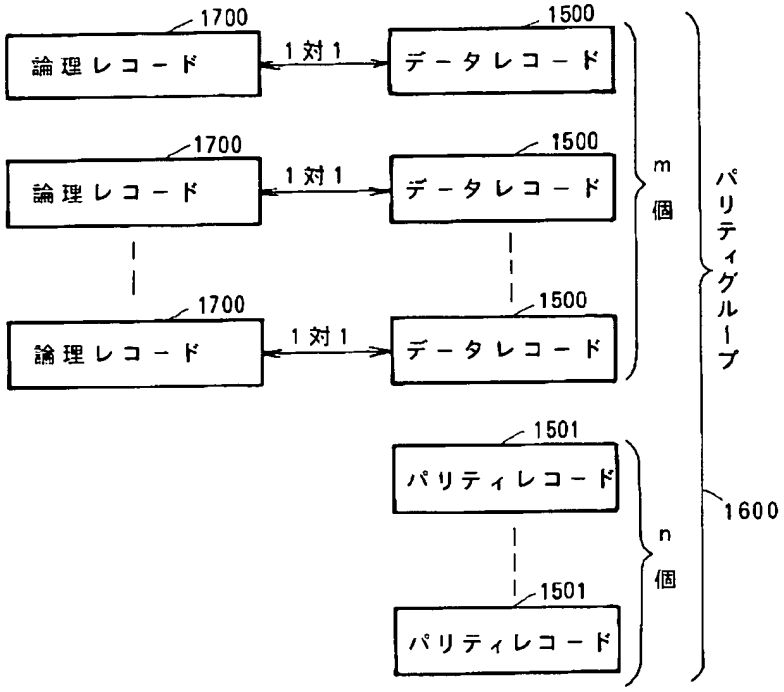
【図３０】

図３０



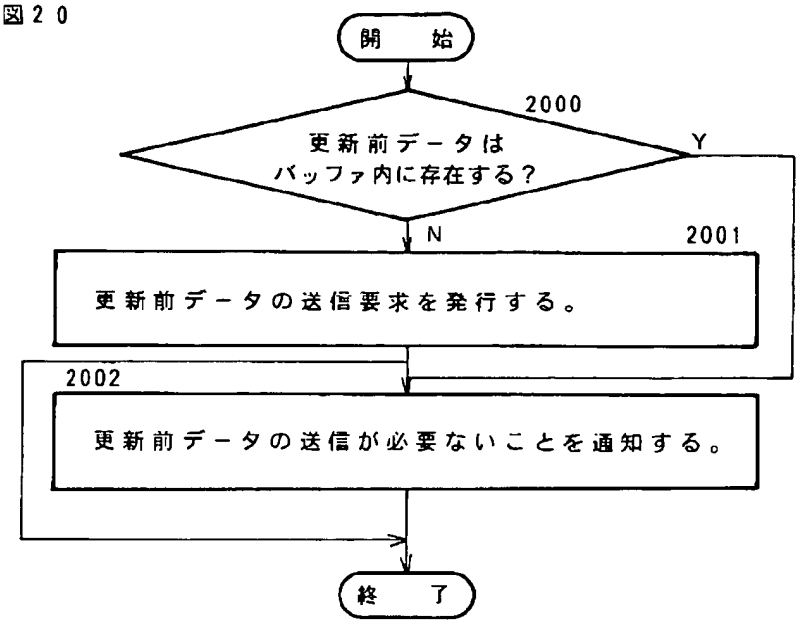
【図１８】

図１８


1700
論理レコード
1対1
1500
データレコード
1700
論理レコード
1対1
1500
データレコード
1700
論理レコード
1対1
1500
データレコード
m
個
1501
パリティレコード
1501
パリティレコード
n
個
パリティグループ
1600

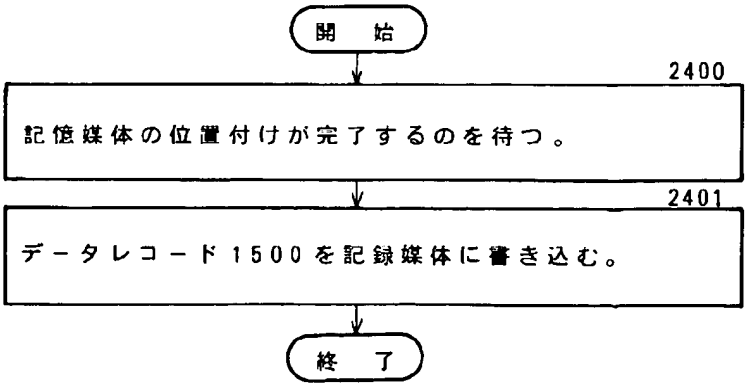
【図２０】

図２０



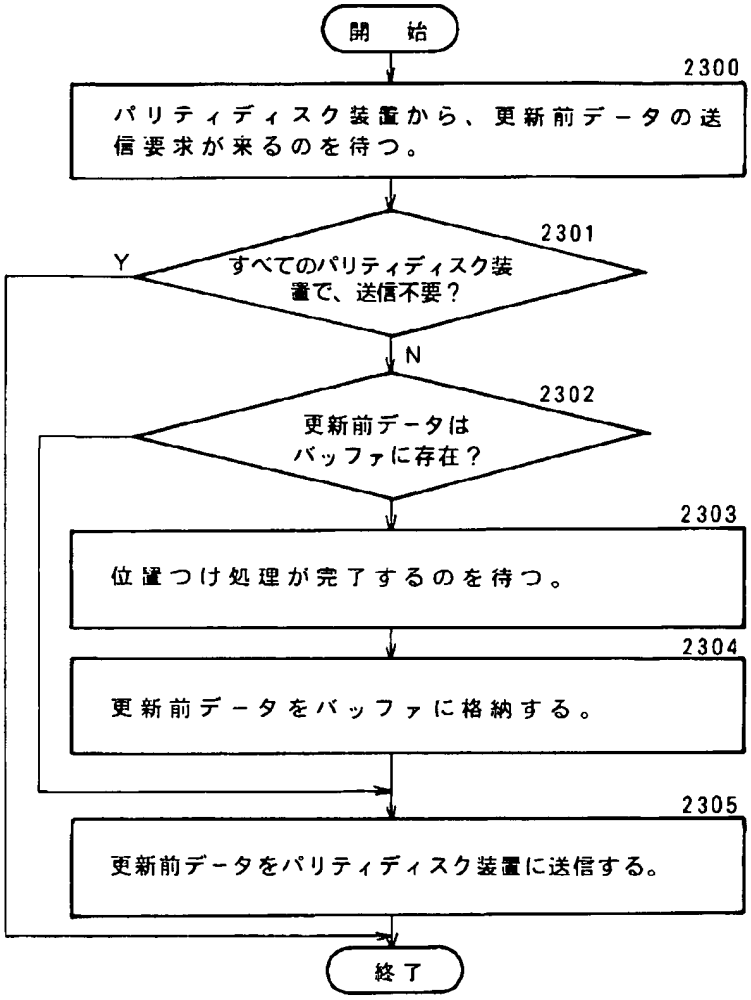
【図２４】

図２４



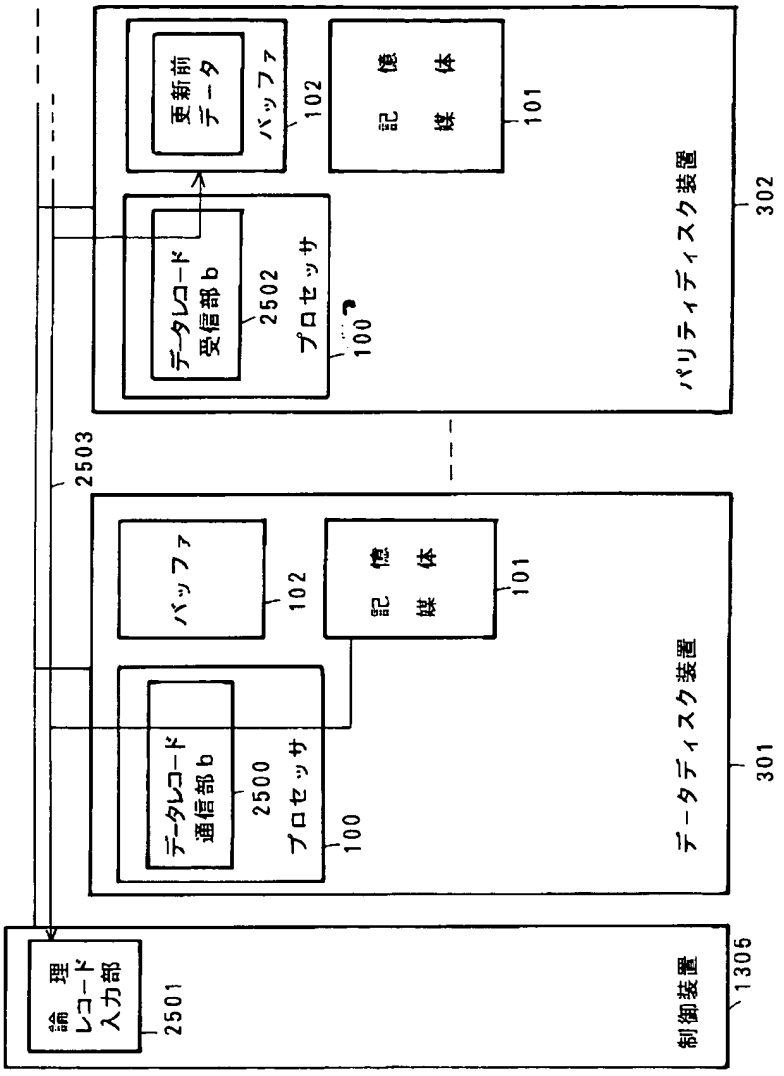
【図２３】

図２３


開　始
2300
パリティディスク装置から、更新前データの送信要求が来るのを待つ。
2301
すべてのパリティディスク装置で、送信不要？
Y
N
2302
更新前データはバッファに存在？
2303
位置つけ処理が完了するのを待つ。
2304
更新前データをバッファに格納する。
2305
更新前データをパリティディスク装置に送信する。
終了

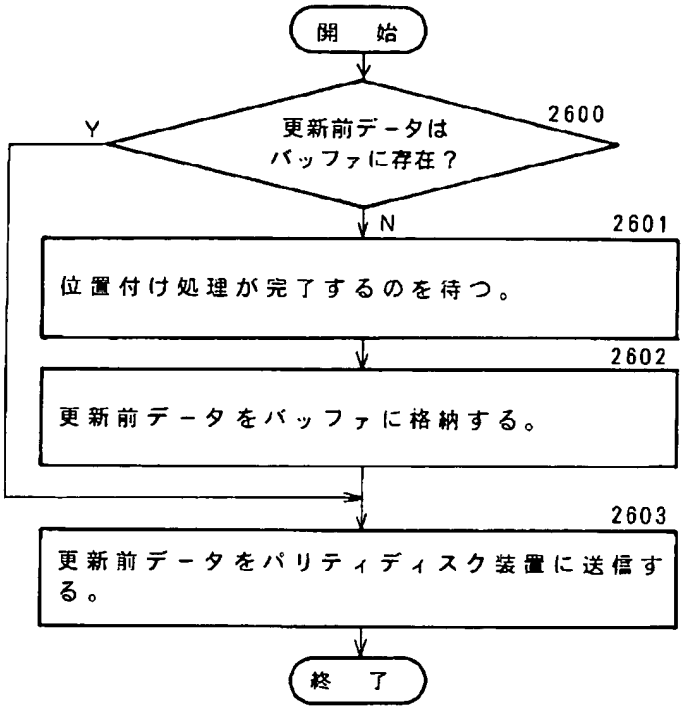
【図２５】

図２５


論理レコード入力部
2501
制御装置
1305
データレコード通信部b
2500
プロセッサ
100
バッファ
102
記憶媒体
101
データディスク装置
301
2503
データレコード受信部b
2502
プロセッサ
100
更新前データ
バッファ
102
記憶媒体
101
パリティディスク装置
302

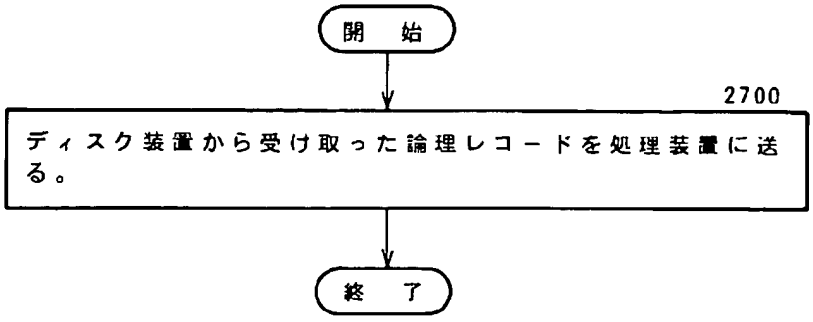
【図２６】

図２６



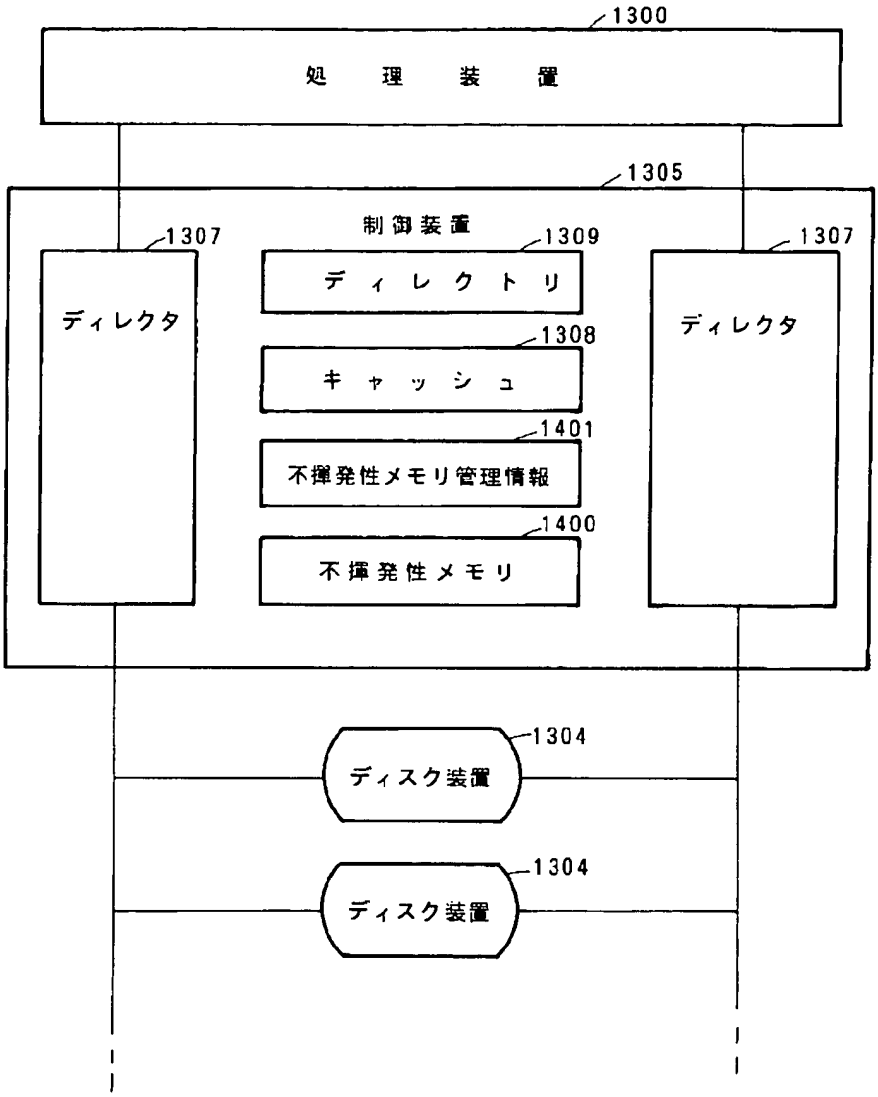
【図２７】

図２７



【図２８】

図２８



フロントページの続き

(72)発明者　倉野　昭
神奈川県小田原市国府津2880番地　株式会社日立製作所小田原工場内

(72)発明者　猪股　宏文
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内